Федеральное государственное автономное образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Институт космических и информационных технологий
Базовая кафедра интеллектуальных систем управления

УТВЕРЖДАЮ
Заведующий кафедрой
_____ Ю. Ю. Якунин
«11» июня 2018 г.

БАКАЛАВРСКАЯ РАБОТА

27.03.03 Системный анализ и управление

Разработка алгоритма генерации случайных чисел

Руководитель _____ _____ Е. А. Чжан
подпись, дата должность, ученая степень инициалы, фамилия

Выпускник _____ А. С. Тихонов
подпись, дата инициалы, фамилия

Красноярск 2018

РЕФЕРАТ

Бакалаврская работа по теме «Разработка алгоритма генерации случайных чисел» содержит 62 страницы текстового документа, 45 рисунков, 4 таблицы, 24 формулы и 29 использованных источников.

КЛЮЧЕВЫЕ СЛОВА: СЛУЧАЙНЫЕ ЧИСЛА, ВЫБОРКА, ЗАКОН РАСПРЕДЕЛЕНИЯ, АЛГОРИТМ ГЕНЕРАЦИИ, ГЕНЕРАЦИЯ ЧИСЕЛ, ИДЕНТИФИКАЦИЯ.

Цель работы состоит в разработке алгоритма генерации выборки псевдослучайных чисел, распределённых по заданному закону распределения.

Для достижения данной цели были поставлены следующие задачи:

– провести анализ существующих алгоритмов псевдослучайных чисел;

– разработать алгоритм генерации псевдослучайных чисел по заданным законам распределения;

– реализовать алгоритм генерации в виде программного модуля;

– провести численные исследования.

Для решения поставленных задач был проведен анализ распространённого программного обеспечения и разработан и реализован алгоритм, который имеет преимущества перед существующими.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Идентификация – совокупность методов для построения математических моделей по данным наблюдений.

Метод статистического моделирования – получение статистических данных об объекте или процессе, которые проходят в моделируемой системе, при помощи средств ЭВМ [1].

Этапы статистического моделирования:

- Генерация псевдослучайного ряда.
- Применение ряда в имитационных моделях.
- Обработка полученных результатов.

Во время процесса статистического моделирования большая часть вычислительной мощности ЭВМ расходуется на генерацию псевдослучайного ряда, поэтому задача упрощения генерации является очень важной.

Актуальность данной работы заключается в том, что случайные числа в наше время используется во многих сферах, от казино до криптографии

Объект исследования в данном работе – это случайна величина, распределенная по заданному закону распределения.

Предмет исследования – Алгоритмы генерации псевдослучайных чисел по различным законам.

Цель данной работы – получить выборку псевдослучайных чисел по заданному закону распределения.

Необходимо решить такие задачи исследования как:

- Провести анализ существующих алгоритмов псевдослучайных чисел.
- Разработать алгоритм генерации псевдослучайных чисел по заданным законам распределения.
- Реализовать алгоритм генерации в виде программного модуля.
- Провести численные исследования.

В данной работе имеется три главы и десять подпунктов.

В первой главе разбирается понятие моделирования, идентификации, постановка задачи идентификации, общие сведения о методах статистического моделирования, существующих алгоритмах генерации случайных чисел и примеры применения алгоритмов генерации в принципе.

Во второй главе описана постановка задачи генерирования, реализованные законы распределения в программном модуле. Так же описан сам алгоритм прецизионного генератора случайных чисел и описано решение специфичной проблемы округления, которая может возникнуть в процессе генерации.

Третья глава содержит вычислительный эксперимент, который позволяет оценить работу генератора при помощи варьирования значений параметров законов распределения и полученных результатов генерации. После эксперимента описано руководство пользователя.

Список источников состоит из 29 пунктов.

1 Идентификация систем

Важную роль в самых различных направлениях науки и техники играет математическое моделирование, так как оно позволяет понимать внутреннее устройство и характеристики тех или иных объектов, процессов и явлений. С развитием ЭВМ стало возможным практическая реализация методов математического моделирования в самых различных сферах, таких как: экономика, психология, медицина, биология и многие другие. Так же, столь разнообразное применение можно объяснить и общностью этих методов [2].

Задачи управления – одна из главных ролей математического моделирования, с помощью него осуществляется, к примеру, анализ системы управления и ее синтез [3].

Для математического моделирования всегда необходимо иметь математическую модель самого объекта или процесса, которая описывает этот самый объект или процесс, но в настоящее время часто случается так, что отсутствует математическое описание необходимого объекта или процесса, поэтому первым шагом математического моделирования является как раз построение математической модели или разработка алгоритма, который позволит получить модель автоматически [4]. По этой причине стали популярны методы построения математических моделей, один из них – идентификация. Она применяется во множестве сфер и тем, таких как [5]:

- принятие решений;
- робастные системы;
- нейронные сети;
- интеллектуальные системы.

Моделью можно назвать объект-копию реального объекта-оригинала, модель заменяет оригинал в определенной окружающей среде для того, чтобы показать свойства и характеристики оригинала. Модели представляют реально существующие процессы, объекты, явления и дают возможность получить

ответы на определенные вопросы [6]. Существует несколько видов моделирования.

Детерминированное и стохастическое моделирование. При детерминированном моделировании каждому набору параметров соответствует определенный набор входных параметров, а при стохастическом моделировании некоторые входные или выходные параметры представляют собой случайные величины [7].

Статическое, динамическое и дискретное моделирование. При статическом моделировании описывается состояние объекта в фиксированный момент времени, а при динамическом моделировании исследуется объект во времени (в основе статистического моделирования лежит метод Монте-Карло). При дискретном моделировании описывается поведение системы только в дискретные моменты времени [8].

Непрерывное и дискретно-непрерывное моделирование. При непрерывном моделировании отображаются непрерывные процессы в системах, а при дискретно-непрерывном моделировании выделяют наличие как дискретных, так и непрерывных процессов.

Мысленное и реальное моделирование. Мысленное моделирование используется, когда нет возможности для физического отображения и когда объекты нереализуемы в заданном временном интервале, но мысленное моделирование может быть реализовано в виде:

– наглядного;

– схематического;

– математического.

При реальном моделировании существует возможность исследования различных характеристик либо на целом реальном объекте, либо на какой-то его части. К видам реального моделирования относятся натурное и физическое моделирование.

Выполнение поиска – характерная черта процедуры идентификации. Поиск подходящей модельной структуры, представительной модели в пределах

этой структуры и т.д. Такая процедура несет в себе итеративный характер – перед тем, как будет сформирована подходящая модель, происходит перебор моделей с отбрасыванием неприемлемых. На данный момент подобная процедура сложно поддается автоматизации, так как решения, которые принимает человек, переплетаются с некоторыми выкладками формального характера и численными расчетами.

Для пакета прикладных программ процесса идентификации можно перечислить несколько типовых составляющих:

– фильтрация данных и их обработка;

– методы идентификации непараметрического характера;

– методы идентификации параметрического характера, предназначенных для модельных структур;

– отображение свойств модели. Имитационное моделирование;

– процессы подтверждения.

Идентификация – это нахождение параметров и построение структуры модели, которая бы позволила обеспечить сходство с самим объектом. Модель строится исходя из данных, которые были получены при нормальном функционировании необходимого объекта или процесса.

Для каждой системы метод идентификации свой. Как таковой классификации нет.

Методы идентификации можно разделить по способу тестирования (активные, т.е. проводиться исключительно для решения задачи, либо пассивные: идентификация происходит, когда система функционирует нормально) и по характеру сигналов (статистические и детерминированные).

1.1 Постановка задачи идентификации

Построение математической модели, которая описывает исследуемый объект или процесс как можно лучшим образом, учитывая выбранный критерий – это и есть задача идентификации [9]. Так же можно интерпретировать это как

нахождение оператора модели, который преобразует данные на входе в величины на выходе. Постановки задачи идентификации могут быть самыми различными, и соответственно, операторы могут быть представлены различными характеристиками и структурой [10].

Вернемся к задаче моделирования и ее все возрастающей важности для практики. Если модель была построена адекватно физическому процессу, то ее можно использовать для численных экспериментов, связанных с изучением поведения процесса или объекта при изменении входных данных, а также влияния внешних воздействий. Так же крайне важно изучение математических и физических связей внутри самого объекта, нахождение закономерностей и меры взаимного влияния параметров и переменных самого процесса, восполнение неточных или потерянных данных [11]. И все описанные выше задачи чаще всего изучить на реальном объекте крайне трудно либо невозможно вовсе [12]. Трудность составляет и затраты времени, и материальные затраты, и остановка либо изменение хода процесса. В этом случае возрастает значимость задачи имитационного моделирования. Модель позволяет реализовать все то, что необходимо для анализа реального объекта, не трогая сам объект или процесс [13].

Ниже, на рисунке 1, представлена общая схема процесса, которая принята в теории идентификации [14].

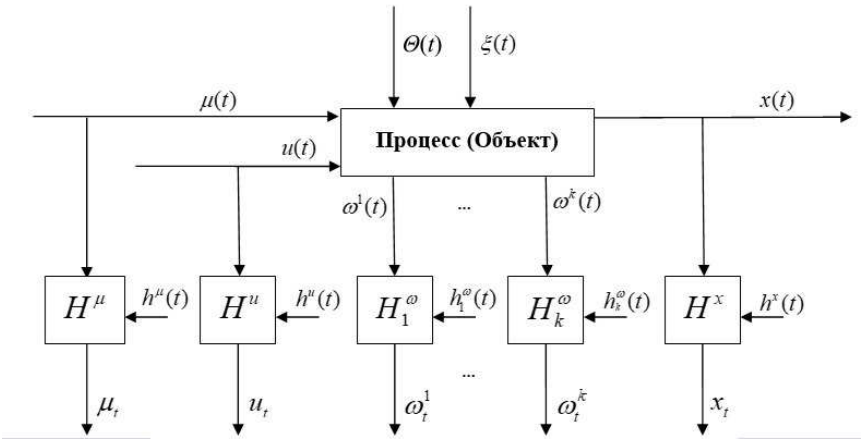

Рисунок 1 – Общая схема процесса идентификации

Обозначения на рисунке: $u(t)$ - векторное управляющее воздействие, $\theta(t)$ – неконтролируемая, не измеряемая переменная, $x(t)$ - векторная выходная переменная процесса, $\omega^i(t)$: $i=1,2,...,k$ - переменные процесса, контролируемые по длине объекта, $\mu(t)$ - векторная входная переменная процесса, $\xi(t)$ - векторное случайное воздействие, t заключенное в круглые скобки – непрерывное время, H со знаком сверху – каналы связи, соответствующие различным переменным, включающие в себя средства контроля, приборы для измерения наблюдаемых переменных, значок t внизу переменных (x,ω,u,μ) - дискретное время [15].

Контролирование переменных (x,ω,u,μ) осуществляется при помощи интервала времени Δt, т.е. $x_i,\omega_i^1,...,\omega_i^k,u_i,\mu_i$, $i=\overline{1,s}$ – выборка измерений переменных процесса $(x_1,\omega_1^1,...,\omega_1^k,u_1,\mu_1)$, $(x_2,\omega_2^1,...,\omega_2^k,u_2,\mu_2),...,(x_s,\omega_s^1,...,\omega_s^k,u_s,\mu_s),...$, s - объем выборки, $h(t)$ - со значком вверху – случайные помехи измерений соответствующих переменных процесса.

Далее рассмотрим этапы, на которые можно разделить задачу идентификации.

1. Обозначить необходимые требования к данным, полученных из наблюдения: перед проведением самого эксперимента нужно обозначить условия как будет осуществляться сбор данных и их последующее использование. Исходя из требуемой задачи и ее точности определяется и количество опытов эксперимента и условий, касающихся его проведения.

2. Исходя из информации об объекте или процессе, которая имеется, нужно определить класс моделей, из которых далее выбирается самая подходящая. Определяется выбор структуры модели и класс уравнений, которые будут описывать процесс. Этот этап часто называют идентификацией в широком смысле. Для приемлемого решения задачи необходимы априорные данные о явлениях, протекающих в процессе, будто то физические, химические или другие [16].

3. Основываясь на задаче, определяются критерии близости модели и исследуемого объекта, но стоить помнить о том, что близость очень

относительна исходя из того, что операторы и модели и объекта иметь неидентичную структуру или число входов, и поэтому адекватность формулируется различными способами. Оценить близость операторов порой очень сложно или вообще невозможно, поэтому можно применять оценивание близости только величин выхода модели и объекта. Чтобы это сделать применяется функция потерь или риска, которая далее подвергается минимизации. Далее исходя из функции потерь или риска определяется критерий, из-за которого задача идентификации далее превращается в задачу оптимизации критерия.

4. Когда структура модели определена, нужно получить численные значения параметров. И в итоге необходимо решить задачу оценивания параметров исходя из данных, полученных экспериментальным путем. Этот этап часто называют идентификацией в узком смысле.

Перед использованием модели необходимо оценить насколько она адекватна т.е. получить оценку качества модели. Проверка адекватности основывается на возможности решения задачи, для которой была построена модель, т.к. полная идентичность модели объекту недостижима. Адекватность, зачатую оценивается подачей одинаковых входных воздействий и сравнение значений на выходе у объекта и модели. При оценке следует использовать данные, отличные от тех, что использовались при идентификации объекта.

Решение задач идентификации можно разделить на несколько подходов:

- идентификация по авторегрессионным моделям;
- идентификация на основе характеристик частоты и времени;
- непараметрические методы;
- рекуррентные методы.

1.2 Общие сведения о методах статистического моделирования

Методы моделирования используются повсеместно в очень многих сферах деятельности человека, если не во всех. Особенно важно моделирование в сфере

управления систем, где решение принимается исходя из полученной информации.

Области применения методов моделирования самые различные. Из наиболее основных областей – это системы массового обслуживания, автоматизированные системы, системы проектирования, системы исследования, управление производственными процессами. Моделирование так же применяется для изучения и анализа того, как работает система, как в ней функционируют и взаимодействуют элементы, как сама система взаимодействует с окружающей средой и анализ самой системы как единый элемент.

Моделирование – это метод воспроизведения и исследования определённого фрагмента действительности (предмета, явления, процесса, ситуации) или управления им, основанный на представлении объекта с помощью модели. Модель похожа на свой объект, но она не должна быть копией объекта, иначе теряется смысл в построении модели, а порой модель вообще не может быть копией объекта по каким-либо причинам [17].

Так же стоит помнить о том, что моделирование зачастую используется в совокупности с другими специальными и научными методами, особенно если речь идет об анализе каких-либо глобальных проблем, которые особенны своей многоплановостью, т.е. включающих в себя, по факту, все сферы деятельности человека. В таких случаях моделирование приобретает вид многомодульного построения. Так же оно сохраняет и свои характеристики при создании модели и более узкоспециализированных проблем, например, социальной сферы: образование, услуги, здравоохранение, демография.

В случаях, когда система подвержена случайными возмущениями, применяются модели, которые называются имитационные. Случайные возмущения – это такие воздействия, которые носят случайный, непредсказуемый точно, характер [18]. В имитационных моделях влияние таких случайных возмущений учитывается при помощи использования таких вещей, как вероятностные характеристики случайных процессов, т.е. законов

распределения вероятности, функции корреляции и другие. Так как данные, полученные в результате отображения исследуемого процесса на модели являются реализациями случайного характера, то для вычисления объективных характеристик исследуемого процесса необходимо большое количество воспроизведений и далее, полученные данные должны быть обработаны статистически. С помощью имитационного моделирования анализ и исследование подобных систем, в которых имеются случайные возмущения называется статистическим моделированием [19].

Предельные теоремы теории вероятности – основа теории метода статистического моделирования на ЭВМ. Определенные закономерности случайных величин и явлений зачастую дают возможность прогнозирования поведения событий и величин, и оценивать их характеристики, которые проявляют устойчивость. Гарантия высокого качества статистического моделирования оценивания исходя из большого количества реализаций N – в этом заключается принципиальное значение таких теорем.

Перед использованием статистического моделирования нужно построить систему стохастического характера, выходные параметры которой позволят оценить искомые. Причем независимо от того, будет исследуемый объект стохастическим или детерминированным.

Сам термин «статистическое моделирование» сильно связан с таким понятием как метод Монте-Карло. Метод основан на применении случайных чисел, т.е. значений некой случайной величины с установленным вероятностным распределением [20]. Метод статистического моделирования – это получение статистических данных об объекте или процессе, которые проходят в моделируемой системе, при помощи средств ЭВМ. Статистические данные классифицируются и обрабатываются при помощи математической статистики [21].

Стохастические воздействия необходимо учитывать во время статистического моделирования. Число случайных чисел колеблется в очень широком диапазоне, речь идет о числах, которые используются для построения

статически устойчивой оценки параметров процесса деятельности некоторой системы S во время реализации алгоритма моделирования на ЭВМ. Зависит это от разных характеристик объекта: его класса моделирования, точность, вид оцениваемых параметров и т.д. Во время статистического моделирования большая часть вычислительных средств ЭВМ уходит на работу со случайными числами. А также, результаты, полученные по время статистического моделирования, сильно подчинены качеству начальному ряду случайных чисел. Исходя из этого можно сделать вывод, что допустимость использования моделирования систем ЭВМ на практике, определяется присутствием экономичных и простых средств генерирования последовательности случайных чисел такого качества, которое требуется.

Статистическое моделирование используется в двух областях:

– при решении задач детерминированного характера;

– при исследовании систем стохастического характера.

Главная идея, используемая для решения задач детерминированного характера с помощью статистического моделирования – это замена задачи детерминированного характера равнозначной схемой системы стохастического характера, выходные параметры которой будут совпадать с выводом решения задачи детерминированного характера. Благодаря замене погрешность моделирования падает при увеличении количества реализаций алгоритма моделирования N.

Таким образом, во время статистической обработки частных данных, полученных во время статистического моделирования системы, получается получить информацию о том, как ведет себя реальный объект или процесс в различные моменты времени. Смоделированная система становится статистически устойчивой относительно приобретённых результатов во время моделирования, если число реализаций N вполне большое и эти результаты можно использовать для оценки искомых параметров процесса деятельности системы S.

Вернемся к ранее упомянутому методу Монте-Карло. Название данного метода произошло от города Монте-Карло, который известен своим казино, ведь одним из примитивных приборов для генерации случайных чисел может служить рулетка [22]. Метод Монте-Карло имеет множество различных приложений. Он применяется в следующих областях: в промышленности для моделирования изменчивости производственных процессов; в физике, химии и биологии для моделирования разнообразных явлений; в области игр для моделирования искусственного интеллекта, например, в китайской игре го; в области финансов для оценки производных финансовых инструментов и опционов.

Современный вариант метода сформировался в рамках Манхэттенского проекта, где он применялся для моделирования расстояний, которые могут пройти нейтроны в различных материалах. Идея моделирования на основе генерации набора случайных значений существовала уже в течение некоторого времени, но особое развитие получила при создании атомной бомбы, распространившись затем во многих других областях знаний.

Большим преимуществом метода Монте-Карло является то, что он позволяет учесть в модели элемент случайности и сложность реального мира. Кроме того, метод является робастным по отношению к изменению различных параметров, таких как распределение случайной величины. В его основе лежит закон больших чисел.

Одним из типичных примеров использования метода Монте-Карло являются задачи, в которых необходимо найти математическое ожидание некоторой случайной величины. Для этого нужно сгенерировать набор случайных значений данной величины и найти среднее. Случайная величина обычно характеризуется определенным распределением вероятностей. При решении задач этим методом нужно достигать последовательности значений случайной величины на ЭВМ с заданным распределением. Этот процесс обычно называют моделированием случайной величины. Такое моделирование случайных величин обычно осуществляется при помощи преобразования одного

или более автономных значений некоторой случайной величины, которая равномерно распределена в интервале от нуля до одного.

Статистическое моделирование, включая метод Монте-Карло, включает в себя несколько этапов:

– генерация псевдослучайного ряда с заданным законом распределения и корреляцией, который при каждой реализации моделирует случайные значения на ЭВМ;

– далее, сгенерированные числовые последовательности применяются в имитационных моделях;

– обработка полученных результатов моделирования с помощью статистики.

На практике в основном используется три способа получения случайных чисел.

– Алгоритмический или программный способ. При этом способе последовательности случайных чисел генерируются при помощи алгоритмов и реализующих эти алгоритмы программ на ЭВМ.

– Аппаратный или физический способ. В этом случае генерация происходит с помощью генератора, который является внешним устройством ЭВМ. Это позволяет не нагружать ЭВМ операциями для генерирования чисел. Недостаток этого способа в том, что он не может давать гарантию о качестве последовательности случайных чисел в то время, когда происходит моделирование на ЭВМ. Также одним из важных недостатков такого метода является невозможность повтора последовательности случайных чисел во время моделирования.

– Табличный или файловый способ. Такой способ предполагает то, что случайные числа, сформированные в отдельный файл и помещенные в оперативную память ЭВМ, приведены к виду таблицы. Недостатком такого способа является то, что его крайне не рекомендуется использовать при большом табличном объеме.

Рассмотрим модели, которые включают в себя методы моделирования:

– индустриальная динамика Джея Форрестера и модель Монте-Карло. Такие модели применяются для моделирования процессов в производственно-хозяйственной сфере;

– принятие решений. Используются в ситуациях, когда интересы сторон расходятся. Стороны определяют стратегию, которая, по их мнению, гарантировала бы максимальную выгоду исходя из действий;

– сетевое планирование. Модель, которая учитывает создание сетевого графика, на котором показан весь комплекс взаимосвязанных работ и алгоритм выполнения действий, которые нужны для реализации конкретной цели. Обычно используется для сокращения времени, отведенного на выполнение сложных проектов.

1.3 Алгоритмы генерации случайных чисел

В современном мире широко используются псевдослучайные числа в самых разных приложениях и сферах:

– моделирование;

– криптография;

– численный анализ (методы Монте-Карло, интегрирование);

– программирование;

– принятие решений (в тех случаях, когда из-за детерминированности происходит замедление процесса);

– сфера развлечений (рулетка, карты и т.д.).

Генератор псевдослучайных чисел – это алгоритм, с помощью которого порождается последовательность чисел, элементы которой подчиняются заданному распределению и почти независимы друг от друга. Сами же числа в последовательности, относительно, независимы друг от друга. Так же числа подчиняются какому-либо заданному закону распределения. Чаще всего на практике используется равномерный закон распределения.

Случайная величина из конкретного диапазона – это величина, вероятность появления которого зависит от функции закона распределения.

Генерация случайных значений возможна лишь, если механизмом генерации выступает химические, физические или природные процессы. С помощью вычислительных машин же генерируются псевдослучайные последовательности значений, для которых сохраняется условие, при котором из-за одинаковых наборов начальных значений происходит генерация одинаковых последовательностей [23].

Псевдослучайные числа могут генерироваться с помощью различных алгоритмических методов, рассмотрим некоторые их них.

Метод средних квадратов. Метод заключается в том, что некое число возводится в квадрат, из полученного числа выделяются знаки из середины, которые и образую случайное число. Повторяется это необходимое количество раз. Подобные число вполне считаются псевдослучайными по той причине, что средние числа зависят от тех, что по краям и которые отбрасываются.

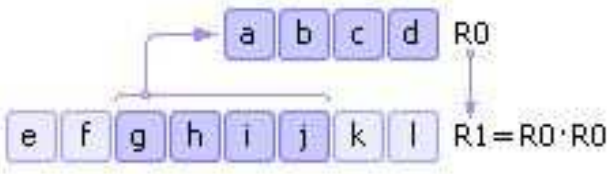

Рисунок 2 – Схема метода средних квадратов

Метод серединных произведений. Метод похож на метод средних квадратов, но только в начале берутся два числа и перемножаются, из полученного числа извлекается середина и умножается на число, на которое было умножено первое.

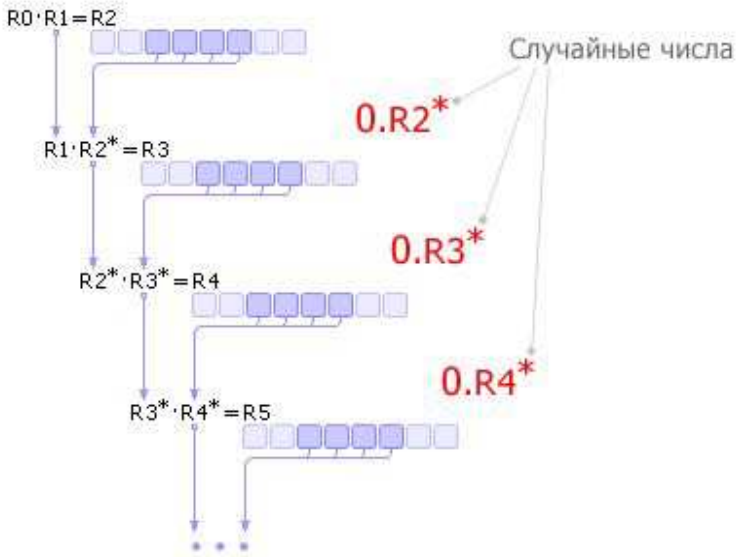

Рисунок 3 – Схема метода серединных произведений

Мультипликативный метод. Суть метода в том, что в начале обозначаются константы C и m, далее берется любое число, а следующее число вычисляется как текущее, умноженное на C и поделенное на модуль числа m.

Метод перемешивания. Идея данного метода основа на том, что берется ячейка, в которой хранится число R0, содержимое этой ячейки сдвигается влево на одну четверть длины самой ячейки, в результате образуется новое число *R*0*, такими же действиями, но со сдвигом вправо образуется число *R*0**. Числа *R*0* и *R*0** складываются и так образуется новое случайное число, далее оно берется как начально и цикл повторяется.

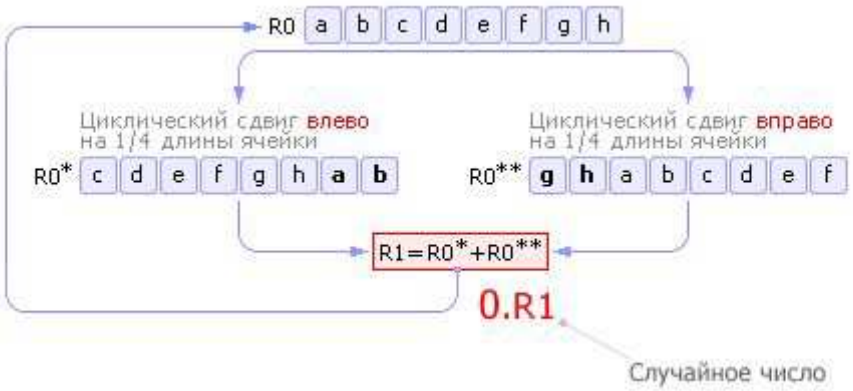

Рисунок 4 – Схема метода перемешивания

Линейный конгруэнтный метод. Случайные числа при помощи этого метода образуются, используя операцию возвращения остатка от деления одного аргумента на другой. Линейный конгруэнтный метод – это один из алгоритмов генерации псевдослучайных чисел. Применяется в простых случаях и не обладает криптографической стойкостью. Почти во всех языках программирования используется этот метод в стандартной функции для получения случайных чисел. Этот метод был впервые предложен Генри Лемером в 1949 году. Выбирается 4 числа:

- модуль m ($m > 0$);
- множитель a ($0 <= a < m$);
- приращение c ($0 <= c < m$);
- начальное значение X_0 ($0 <= X_0 < m$).

Последовательность случайных чисел вычисляется с помощью рекуррентной формулы (1)

$$X_{n+1} = (a * X_n + c)\ mod\ m. \quad (1)$$

Данный метод дает хорошие псевдослучайные числа, но, если брать числа m, a, c произвольно, то результат будет плохим. При $m = 7$, $X_0 = 1$, $a = 2$, $c = 4$ получится последовательность 1, 6, 2, 1, 6, 2, 1, и т.д. Очевидно, что такая последовательность не совсем является случайной, но исходя из результатов можно сделать вывод, что числа m, a, c, X_0 не должны быть случайными и Линейный конгруэнтный метод даёт нам повторяющиеся последовательности.

Как уже было сказано выше, данный метод не является криптографически стойким по той причине, что, зная четыре подряд идущих числа, можно составить систему равнений, с помощью которых можно найти переменные a, c, m.

Аддитивный метод или генератор Фибоначчи. Каждое следующее число высчитывается из суммы двух предыдущих.

Для определения качества генератора последовательности необходимо из псевдослучайных чисел построить гистограмму. При равномерном законе распределения высота столбцов гистограммы должна быть приблизительно равна. Генераторы псевдослучайных чисел обладают и своими недостатками: значения распределяются неравномерно, среди соседних последовательных значений имеется зависимость и довольно короткий период генерируемой последовательности.

1.4 Примеры применения алгоритмов генерации случайных чисел

Случайные числа находят свое применение во многих сферах и задачах программирования, некоторые из них приведены ниже:

- имитационное моделирование;
- численный анализ;
- имитация ввода пользователя;
- тестирование различных алгоритмов.

Генераторы случайных чисел широко используются в веб-безопасности для шифрования, генерации паролей, генераторов сессий и т.д.

Самая используемая в настоящее время сфера применения случайных чисел – это моделирование. Основа моделирования – это модель реального объекта, которая отражает какие-либо черты реального объекта, которые необходимы для исследования. Случайные числа в данном случае используются для воспроизведения явлений из реальной жизни.

Свое применение случайные числа нашли и в криптографии, где генератор используется как криптосистема с секретным ключом, либо используются как модель блочных шифров.

Случайные числа могут быть использованы для построения алгоритмов, которые используют случайные биты. С помощью генератора псевдослучайных чисел можно брать малое количество случайных битов, а другие генерировать, применяя генератор.

Одним из примеров использования метода Монте-Карло является вычисление площади какой-либо фигуры. Фигура помещается в квадрат и на этот самый квадрат случайным образом ставятся точки. Площадь фигуры вычисляется по частоте попадания точек на нее.

Генераторы, основанные на линейном конгруэнтном методе в основном не применимы для криптографии, но находят свое применение, например, в сфере моделирования. При эмпирических тестах показывают отличные статистические характеристики.

Генераторы, основанные на методе Фибоначчи, могут использоваться в статистических алгоритмах, которым необходимо высокое разрешение.

Так же порой возникает необходимость генерирования абсолютно непредсказуемые случайные числа. Их генерация происходит с помощью генераторов случайных чисел. Такие генераторы выдают последовательность, повторить которую невозможно. Одной из главных сфер применения подобных генераторов является создание уникальных ключей шифрования. Такие ключи могут быть сгенерированы и с помощью генераторов псевдослучайных чисел, но в таком случае их безопасность резко падает.

Выводы по первой главе

Одна из главных задач, которая стоит перед математическим моделированием различных химико-технологических процессов – это идентификация, так как именно она является главным этапом для построения адекватной модели процесса.

Идентификация представляет собой список действий, требуемый для построения математической модели.

Методы идентификации систем – это гибкий способ решения большого круга задач в сферах техники и науки. Ее ценность доказана многочисленными приложениями из самых различных сфер и областей. В настоящее время есть подтверждение применения методов идентификации в таких сферах, как:

управление, медико-биологические системы, сейсмология, окружающая среда, экология, эконометрика, обработка сигналов и другие.

Для методов идентификации существуют и ограничения. Основное ограничение – качество данных. Если данные плохие, либо их мало, то от этого сильно страдает процесс идентификации, а порой и вообще нереализуем ввиду слишком малого количества данных. Одна из наиболее часто встречающихся причин малого количества данных состоит в том, что данные с объекта можно снимать через очень большое количество времени. От этого зачастую сильно страдают системы в сфере экономики и экологии.

В программировании последовательности случайных чисел применяются довольно часто для различных задач. Они применяются при тестировании алгоритмов и программ, моделировании процессов, имеющих случайных характер, и отладке. Способ получения случайных чисел может быть с помощью программных, табличных или аппаратных генераторов. Действительно случайные числа программный генератор пока что не может создавать, числа, полученные таким путем, называются псевдослучайными. В языках программирования подобные генераторы имеются в качестве функций по умолчанию. Действительные случайные числа можно получать лишь с помощью аппаратных генераторов.

Моделирование и программирование в целом не могло бы существовать без случайных чисел.

Из вышесказанного можно выразить и главное требование, выдвигаемое последовательностям случайных чисел. Они должны быть случайны и непредсказуемы.

2 Прецензионный генератор псевдослучайных чисел

2.1 Постановка задачи генерирования

Последовательность случайных чисел часто применяется при моделировании для повышения адекватности модели. Сама по себе последовательность случайных чисел – это имитация того, как система работает в реальных условиях.

Задача генерации сводится к формированию множества псевдослучайных чисел, которые распределены по какому-то определенному закону распределения.

Математическая постановка задачи генерирования может быть описана так: требуется сгенерировать выборку значений псевдослучайной величины $s = \{s_1, s_2, \dots, s_n\}$ размера *n*, все элементы которой есть элемент некоторого множества $u = \{u_1, u_2, \dots, u_N\}$ всех возможных значений случайной величины *x,* которая распределена по заданному закону распределения *F(x),* с определённой плотностью распределения *f(x).*

Существует критерий согласованности, который обозначает функционал, характеризующий степень согласия выборки значений С.В. и её закона распределения. В ходе анализа критериев согласованности было отмечено две группы критериев, отличающиеся самим принципом определения степени согласия.

1. Параметрические критерии. Принцип работы таких критериев следующий: исследователь по некоторому правилу (например, используя метод максимального правдоподобия, метод моментов и т.д.) определяет оценки параметров закона распределения. Полученные оценки сравниваются со значениями истинных параметров закона распределения С.В. Разница между оценками параметров и значениями параметров характеризует степень согласованности выборки значений С.В. и закона распределения С.В.

2. Непараметрические критерии. В группу этих критериев входят критерии, рассчитывающие степень согласия при помощи сравнения закона распределения С.В. с эмпирическим законом распределения С.В. К таким критериям относятся: критерии Колмогорова-Смирнова, Пирсона, Андерсона-Дарлинга, а также ряд критериев, специализированных для нормального распределения: *Z*-тест, критерии Жака-Бера, Шапиро-Уилко.

2.2 Реализованные законы распределения

В данном программном модуле реализовано четыре закона распределения, с помощью которых осуществляется генерация случайной величины:

- нормальный закон распределения;
- закон распределения Лапласа;
- закон распределения Парето;
- экспоненциальный закон распределения.

Нормальный закон распределения, который часто называют законом Гаусса, занимает важное место в теории вероятностей и чаще других законов встречается на практике.

Важнейшей чертой, которая выделяет этот закон распределения среди других, заключается в том, что он представляет собой предельный закон, к которому приближаются остальные законы распределения при довольно часто встречающихся типичных условиях. Функция плотности нормального распределения представлена на рисунке 5, где на оси абсцисс располагается значение случайной величины, а на оси ординат значение функции плотности.

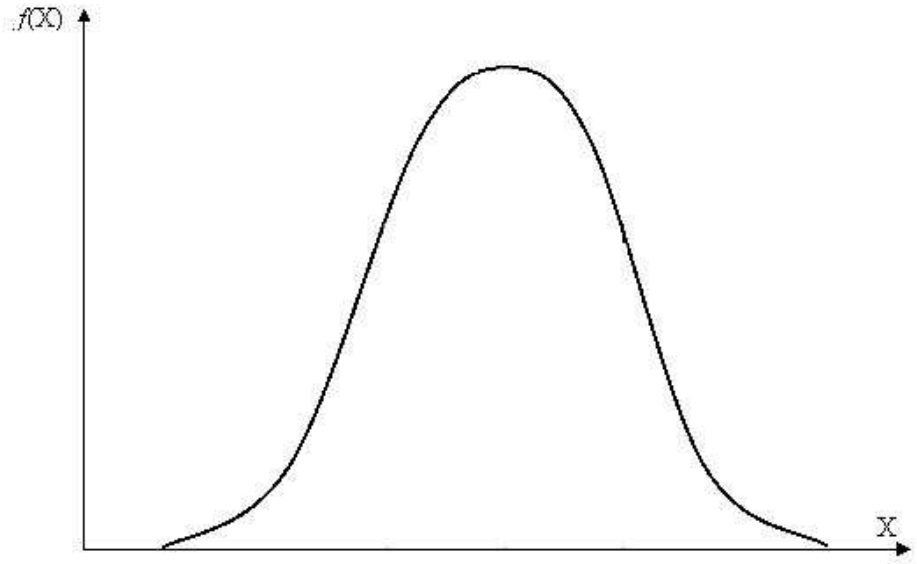

Рисунок 5 – Функция плотности нормального распределения

Кривая по своей форме похожа на колокол, именно поэтому график часто называют колоколообразной кривой. Суть нормального закона распределения заключается в «горбе» графика в середине и резкий спад плотности по краям. Иначе говоря, вероятность того, что случайная величины попадет в районе центра намного выше, чем то, что она бы отклонилась от середины. Дисперсия из исходной формулы закона распределения (2) определяет остроконечность кривой. Весь массив данных сконцентрирован у центра, если они имеют небольшой разброс. В обратном случае, данные «расплывутся» на широкий диапазон, если их разброс большой [24].

В качестве примера, где встречается нормальное распределение, можно привести природу. Это могут быть характеристики живых организмов, состоящих в популяции, например, рост, вес. Так же это могут быть физические и умственные способности людей и т.д. По каждому признаку имеется основная масса и отклонения в обе стороны от основной массы. Отклонения при стрельбе так же является примером нормального распределения.

Широкое распространение нормального закона распределения обусловлено тем, что математическое описание закона является бесконечно делимым непрерывным распределением с конечной дисперсией. Именно

поэтому к нормальному распределению в пределе приближаются другие законы, такие как пуассоновское и биномиальное. Нормальное распределение зачастую применяется при моделировании недетерминированных физических процессов. Так же следует упомянуть про применение нормального распределения в изучении параметров психиатрии и психологии человека. В таком случае, из-за анализа многомерных случайных величин, используется многомерное нормальное распределение.

Закон распределения Лапласа, которое так же является двойным экспоненциальным, по виду кривой похож на нормальное распределение, кривая так же симметрична, относительно математического ожидания и коэффициент асимметрии равен нулю, но кривая распределения Лапласа является островершинной, что значит, что у нее тяжелые «хвосты» и высокий пик. Это видно на рисунке 6.

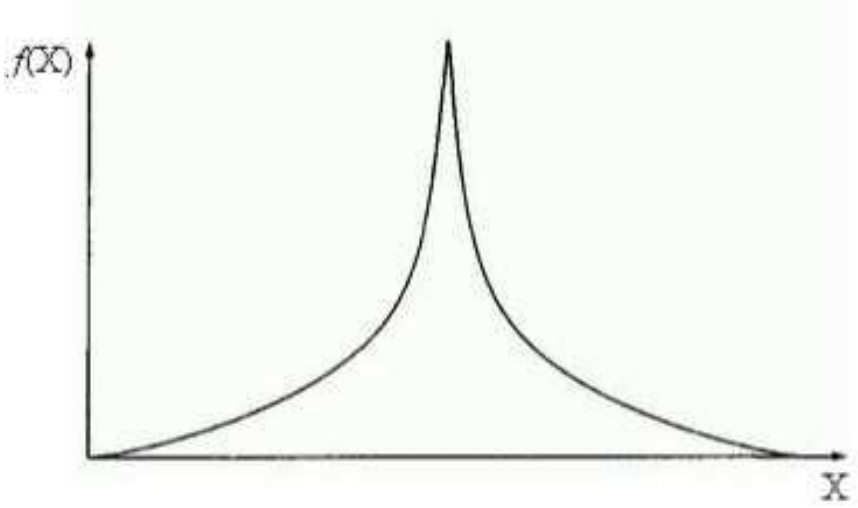

Рисунок 6 – Функция плотности распределения Лапласа

Распределение Лапласа может применяться для описания логарифмов относительного изменения цен активов, обычно с большим успехом, чем нормальное распределение. Так же данное распределение может применятся при моделировании обработки сигналов или в моделировании биологических процессов, финансах и экономике, а также применяться для кредитных рисков и страховых случаях.

Закон распределения Парето в теории вероятностей – двухпараметрическое семейство абсолютно непрерывных распределений, являющихся степенными. Кривая, в отличии от распределения Парето и нормального распределения, не симметрична, а является усеченной, где основная масса данных сосредоточена в «хвосте».

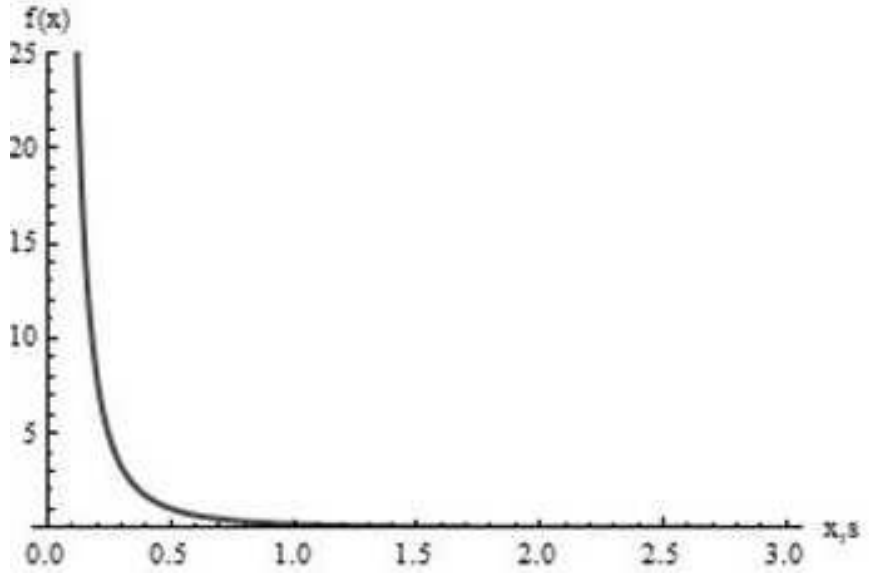

Рисунок 7 – Функция плотности распределения Парето

Чаще всего встречается при изучении разного рода явлений, таких как: социальные, физические, экономические и другие. Изначально использовалось при описании распределения благосостояния и распределения дохода. «Правило 20 к 80», которое означает, что 20% населения владеет 80% богатств, зависит от конкретной величины k, и утверждается, что фактически встречаются существенные количественные отклонения, например, данные самого Парето по Британии говорят, что там примерно 30% населения владеет 70% общего дохода.

В лингвистике распределение Парето известно под именем закона Ципфа (для разных языков показатель степени может несколько различаться, также существует небольшое отклонение от простой степенной зависимости у самых частотных слов, однако в целом степенной закон описывает это распределение достаточно хорошо). Частным проявлением этой закономерности

можно считать зависимость абсолютной частоты слов (сколько всего раз каждое конкретное слово встретилось) в достаточно длинном тексте от ранга (порядкового номера при упорядочении слов по абсолютной частоте). Степенной характер остается вне зависимости от того, приводятся ли слова к начальной форме или берутся из текста как есть. Аналогичная кривая для популярности имен и Распределение размера населенных пунктов.

Экспоненциальный закон распределения является непрерывным, который моделирует время между двумя последовательными исполнениями одного и того же события. Так же экспоненциальное распределение – это частный случай распределения Вейбулла. Кривая функции плотности экспоненциального распределения так же, как и кривая распределения Парето не является симметричной.

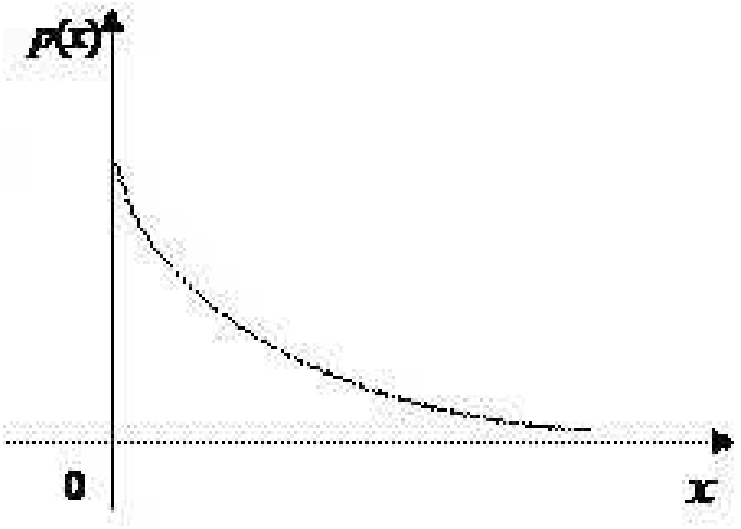

Рисунок 8 – Функция плотности экспоненциального распределения

Так как экспоненциальное распределение позволяет моделировать интервалы времени между наступлением событий, оно играет важную роль в задачах в сфере телекоммуникации. Экспоненциальный закон распределения занимает особое место в теории надежности и теории массового обслуживания [25].

Следует так же заметить, что из экспоненциальных величин строятся другие важные величины, которые, например, имеют другое распределение, такое как Эрланга. Экспоненциальный закон распределения применяется и в оценке длительности срока службы технических устройств. В частности, этому закону распределения следует и наработка между отказами ремонтируемых изделий при их работе на установившихся режимах, и время наработки до отказа некоторых неремонтируемых изделий. Для отказа сложных систем, которые состоят из однотипных деталей, распределение также характерно [26].

2.3 Алгоритм П-генератора случайных чисел

Основанием для создания прецизионного генератора псевдослучайных чисел послужил принцип, на котором основываются непараметрические критерии согласия, а именно принцип согласованности эмпирического распределения случайной величины и закона распределения случайной величины. Правило работы генератора состоит в генерации чисел, которые не зависят друг от друга и являются сами по себе реализациями некоторой случайной величины, для которой закон распределения – это оценка желаемого закона распределения.

Метод статистического моделирования уже давно применяется в самых разных задачах кибернетики для изучения алгоритмов идентификации, управления, распознавания образов и т.д. Новые формулировки задач не совсем поддаются строгой математической постановке и в связи с этим в последнее время появилось много эвристических алгоритмов, что указывает на отсутствие процедуры аналитического синтеза различных алгоритмов, доказательства теорем сходимости, которые подтверждали правомерность дальнейших действий. Исходя из этого метод статистического моделирования необходимо считать этапом доказательства, а никак не иллюстрацией работы алгоритмов. Последнее сильно повышает требования для осуществления такого исследования. Главное в данном случае – это возможность для других

исследователей повторить определенный цикл численных экспериментов. Здесь важный аспект – необходимость работы со случайными помехами, которые распределены по конкретному закону [27]. На практике те генераторы случайных чисел, которые существуют на данный момент [28], лишь условно можно назвать подобающими заявленным законам распределения. Отклонение хорошо заметно при небольших объёмах выборок. Алгоритм, которые предложен ниже, решает эту проблему и может применяться для выборок как большого, так и малого объема.

Постановку задачи для нашего случая можно обозначить так: необходимо получить выборку случайно распределенной величины *x* по выбранному закону распределения.

Ниже, на рисунке 9, приведена блок-схема алгоритма работы генератора.

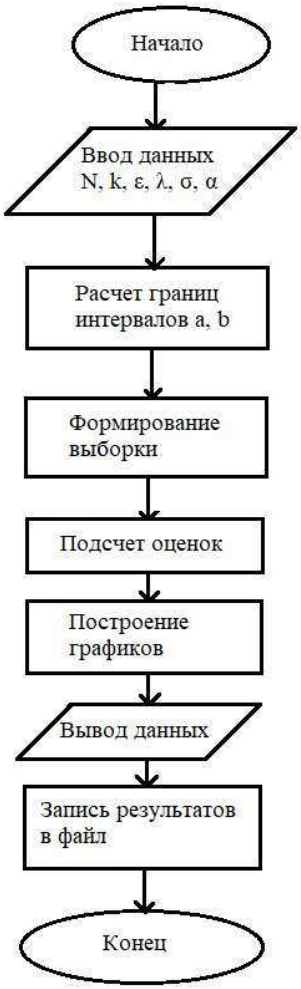

Рисунок 9 – Блок схема генератора псевдослучайных чисел

Параметры:

k – количество интервалов. Определяет количество отрезков, на которых будут разбрасываться точки при построении заданного закона;

N – объем выборки;

ε – точность. Под точностью понимаем наименьшее значение плотности распределения Лапласа. С помощью этого значения находим границы интервала a, b;

λ – параметр масштаба закона распределения Лапласа и экспоненциального закона распределения;

σ – параметр масштаба нормального закона распределения;

α – параметр масштаба закона распределения Парето.

Алгоритм работы генератора состоит из нескольких этапов:

1. Задаются такие параметры как: объем выборки, количество подынтервалов k, точность e, необходимый закон распределения и значения параметров закона распределения.

2. Определяются границы [a, b] генерирования случайной величины x. Границы для каждого из законов определяются по-разному, исходя из исходной формулы закона распределения. Для начала приведем формулы для вычисления плотностей законов распределения:

Нормальное распределение представлено на формуле (2)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-m_x)^2}{2\sigma^2}}, \quad (2)$$

где σ – среднеквадратичное отклонение величины x;

m_x – математическое ожидание величины x.

Распределение Лапласа представлено на формуле (3)

$$f(x) = \frac{\lambda}{2} e^{-\lambda|x|}, \quad (3)$$

где λ - параметр масштаба;

Распределение Парето представлено на формуле (4)

$$f(x) = \frac{a}{C_0}\left(\frac{C_0}{x}\right)^{a+1}, \quad (4)$$

где a – параметр масштаба;

C_0 – начальная граница.

Экспоненциальное распределение представлено на формуле (5)

$$f(x) = \lambda e^{-\lambda}, \quad (5)$$

где λ – параметр масштаба.

3. Далее из приведенных выше формул определяются границы [a, b]. Формула плотности вероятности закона распределения приравнивается к ε и вычисляется значение x для законов распределения Лапласа и нормального значения получаются два, большее – левая граница, меньшее – правая граница. Для остальных законов вычисляется лишь одно значение – правая граница. Левая для других законов либо задается пользователем, как для закона Парето, либо равна нулю, как для экспоненциального закона распределения [29].

Для нормального распределения формулы (6) и (7)

$$a = -\sqrt{-2\sigma^2 \cdot \ln(\varepsilon\sigma\sqrt{2\pi})} + m_x, \quad (6)$$

$$b = \sqrt{-2\sigma^2 \cdot \ln(\varepsilon\sigma\sqrt{2\pi})} + m_x. \quad (7)$$

Для распределения Лапласа формулы (8) и (9)

$$a = -\frac{\ln\left(\frac{\lambda}{2\varepsilon}\right)}{\lambda}, \tag{8}$$

$$b = \frac{\ln\left(\frac{\lambda}{2\varepsilon}\right)}{\lambda}. \tag{9}$$

Для распределения Парето формулы (10) и (11)

$$a = C_0, \tag{10}$$

$$b = \left(\frac{a * {C_0}^a}{\varepsilon}\right)^{\frac{1}{a+1}}. \tag{11}$$

Для экспоненциального распределения формулы (12) и (13)

$$a = 0, \tag{12}$$

$$b = -ln\frac{\varepsilon}{\lambda} * \frac{1}{\lambda}. \tag{13}$$

4. На интервале [a, b] необходимо вычислить подынтервалы, в переделах которых, впоследствии, и будут генерироваться случайные величины. Вычисление границ подынтервалов производится по следующей формуле (14)

$$l_i = l_{i-1} + \Delta, \quad i = \overline{2,k}, \tag{14}$$

где значение Δ определяется по формуле (15)

$$\Delta = \frac{b-a}{k}. \tag{15}$$

5. После того, как были вычислены подынтервалы, необходимо определить количество точек генерирования для каждого из них. Применим закон больших чисел, который устанавливает связь между вероятностью события и его частотой

$$p_i = \frac{n_i}{N}, \quad i = \overline{1,k}, \tag{16}$$

где n_i – число точек в i-м подынтервале.

Из формулы (16) получаем формулу, которая обозначает количество точек, которые необходимо сгенерировать в подынтервале

$$n_i = p_i \cdot N, \quad i = \overline{1,k} \tag{17}$$

Неизвестную, в данном случае, переменную p_i определим по формуле (18)

$$p_j = \frac{f(l_{j-1}) + f(l_j)}{2} * (l_j - l_{j-1}), \quad j = \overline{1,k}. \tag{18}$$

Данный шаг обусловлен тем, что вероятность характеризуется вычислением площади области распределения вероятности. На рисунке 10 это наглядно представлено.

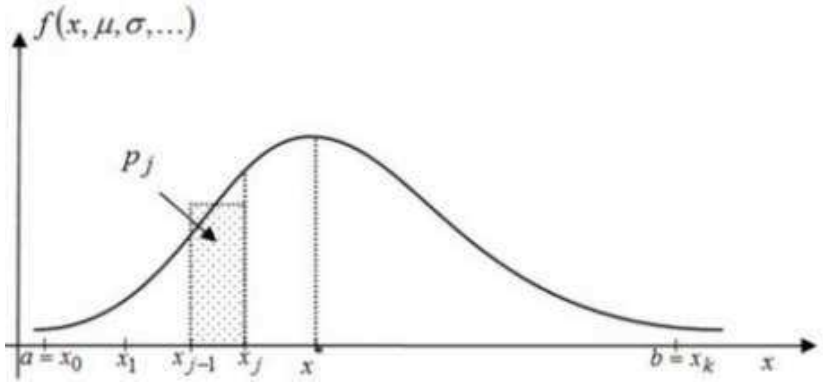

Рисунок 10 – Графическая интерпретация алгоритма

6. Сгенерируем необходимое количество точек n_i в i-м подынтервале, где $i = \overline{1,k}$ по равномерному закону встроенной функцией генерации. Данный шаг

необходим для приемлемого результата работы закона распределения при малом объеме выборок. Это хорошо заметно при сравнении результатов работы нормального закона в стороннем программном обеспечении и в разработанном. Форма «колокола» при малых выборках достигается благодаря тому, что точки в подынтервале генерируются при помощи равномерного закона распределения.

7. Перемешиваем выборку при помощи метода пузырька. Этот этап необходим для того, чтобы можно было использовать сгенерированную выборку для задач имитационного моделирования.

8. Находим оценку параметра выбранного закона распределения по соответствующей формуле для каждого закона. Нормальное распределение (19), распределение Лапласа (20), распределение Парето (21), экспоненциальное распределение (22)

$$\hat{\delta} = \sqrt{\frac{1}{N-1} * \sum_{i=1}^{N} x_i^2}, \tag{19}$$

$$\hat{\lambda} = \sqrt{\frac{2}{\frac{1}{N-1} * \sum_{i=1}^{N} x_i^2}}, \tag{20}$$

$$\hat{a} = \frac{m_{x_i}}{m_{x_i} - C_0}, \tag{21}$$

$$\hat{\lambda} = \frac{N-1}{\sum_{i=1}^{N} x_i}. \tag{22}$$

9. На данном этапе происходит построение графика и гистограммы, на которых можно увидеть насколько соответствует оценка плотности вероятности и истинные значения. Гистограмма строится на основе полученных значений случайной величины и плотности вероятности. Количество столбцов зависит от количества заданных интервалов, а высота характеризует вероятность попадания значения случайной величины в конкретный подынтервал. На графике строится

две кривых, одна является графиком истинной плотности распределения, а вторая ее оценка, где в качестве параметра используется оценка, полученная по выборке значений случайной величины. Исходя из соответствия значений полученных оценок заданных параметров можно сделать вывод о точности работы генератора.

2.4 Решение проблемы округления

На шаге алгоритма, когда происходит генерация необходимого количества точек в i-м подынтервале, может возникнуть проблема с количеством точек в подынтервале. Данная проблема решается для каждого закона по-своему. Сначала осуществляется проверка соответствия полученного количества точек и заданного пользователем числа. Далее исходя из закона решается проблема нехватки или избытка точек. Для нормального распределения и распределения Лапласа решить данную проблему можно удалив точки с концов интервала [a, b] при помощи встроенной функции, в случае, если объем выборки превышает N. Если же наоборот, точек в выборке не хватает, следует добавить недостающее количество в середину выборки при помощи встроенной функции *Random.Next*. Для распределения Парето и экспоненциального распределения проблема решается тем, что если количество точек превышает N, то точки удаляются с конца интервала тем же способом, что и в случае с законами распределения Лапласа и нормальным законом, но если точек не хватает, то новые точки, которые рассчитываются как среднее значение из соседних, вставляются равномерно по всему интервалу. Равномерное распределение осуществляется при помощи формулы представленной ниже

$$Z = \frac{N}{C}, \tag{23}$$

где N – заданное количество точек;

C – недостающее количество точек, рассчитываемое по формуле (24)

$$C = N - \sum_{i=1}^{N} X_i. \quad (24)$$

Выводы по второй главе

Сгенерировать выборку случайной величины, которая была бы распределена по конкретному закону распределения – это задача генерации.

Каждый закон распределения уникален по-своему и применяется на практике для определенных задач. Для охвата широкого спектра задач было реализовано четыре закона распределения, которые чаще всего встречаются на практике.

Был описан подробный алгоритм генератора псевдослучайных чисел с учетом всех используемых формул для каждого из законов, а также приведена блок-схема алгоритма.

Избыток или нехватка значений – серьезная проблема. Для каждого закона распределения эта проблема решается своим методом.

3 Численные исследование алгоритма генерации чисел

3.1 Вычислительный эксперимент

Чтобы оценить работу генератора с законами распределений, необходимо провести численные исследования. Для сравнения с результатами программного модуля был реализован нормальный закон распределения в таком программном обеспечении как: Microsoft Excel, MathCad, Matlab. Варьирование осуществлялось только по объему выборки N.

В Microsoft Excel нормальный закон распределения был реализован при помощи встроенной функции «НОРМ.РАСП» и значениях $\sigma = 1$, $m = 0$. Пример использования этой функции представлено на рисунках 11, 12 и 13.

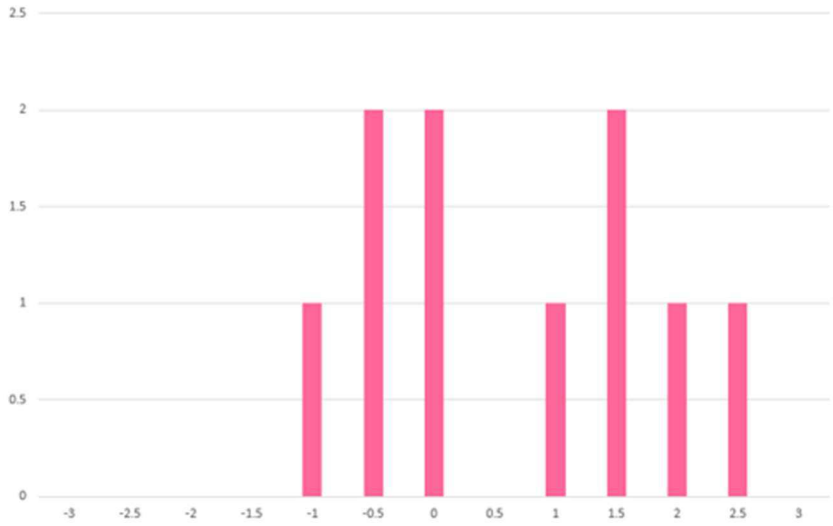

Рисунок 11 – НОРМ.РАСП при $N = 10$

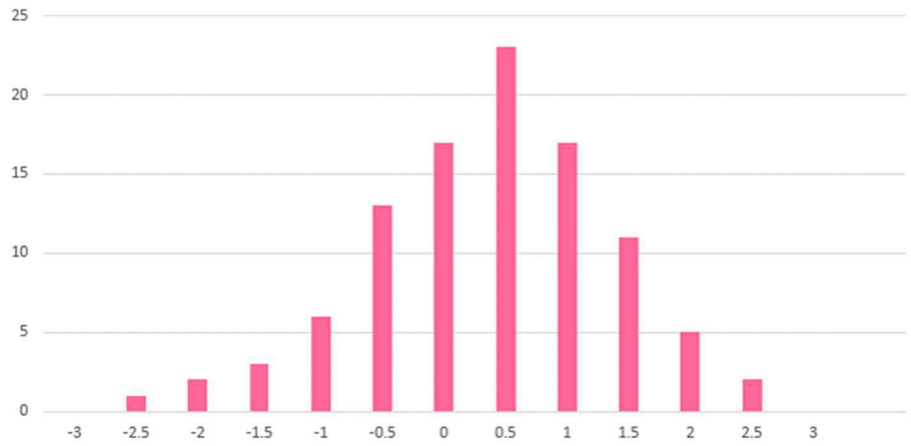

Рисунок 12 – НОРМ.РАСП при $N = 100$

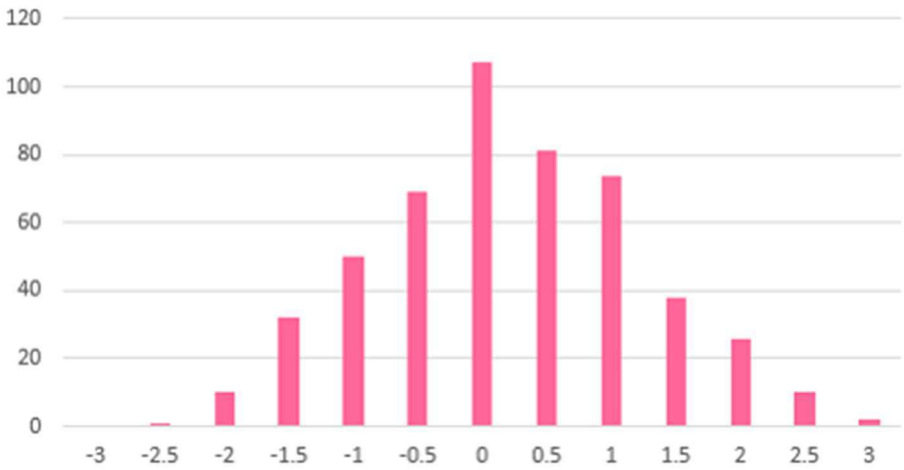

Рисунок 13 – НОРМ.РАСП при $N = 500$

При $N = 10$ оценка математического ожидания и среднеквадратичного отклонения были равны $\hat{m} = 0.42$ и $\hat{\sigma} = 1.215$. При $N = 100$: $\hat{m} = 0.11$ и $\hat{\sigma} = 0.976$. При $N = 500$: $\hat{m} = 0.05$ и $\hat{\sigma} = 1.020$.

В MathCad нормальный закон распределения был реализован при помощи встроенной функции dnorm и значениях $\sigma = 1$, $m = 0$. Пример использования этой функции представлено на рисунках 14, 15 и 16.

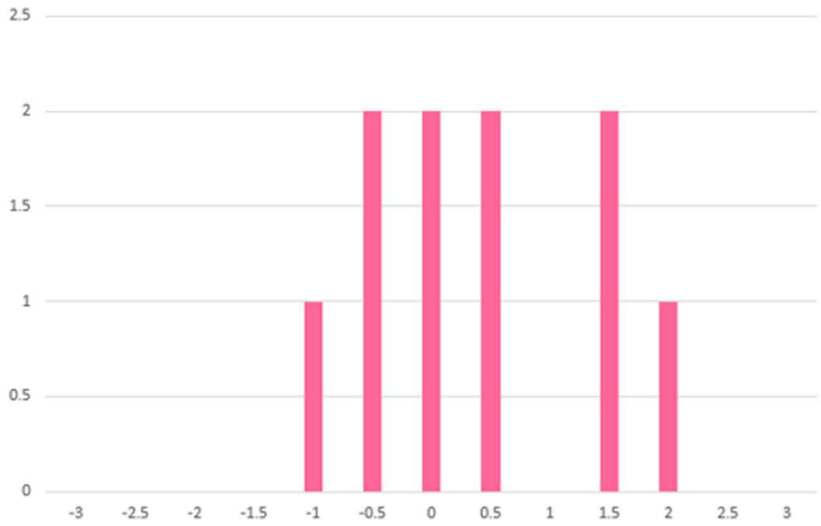

Рисунок 14 – dnorm при $N = 10$

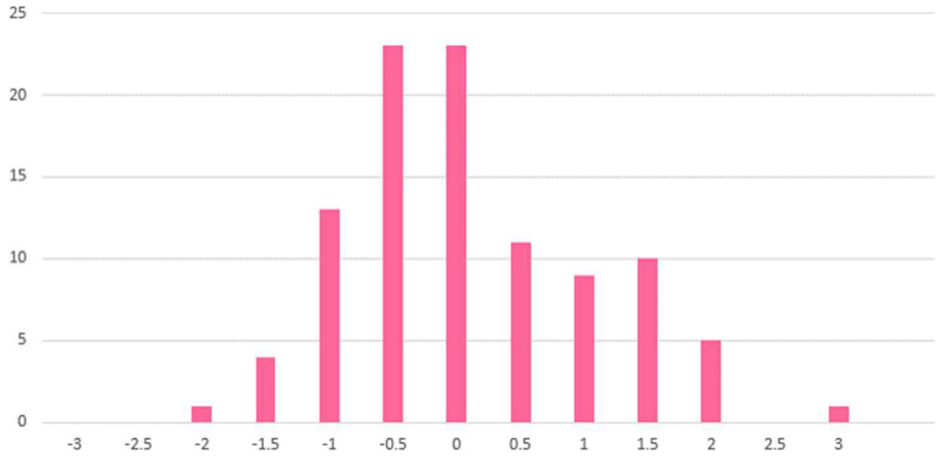

Рисунок 15 – dnorm при $N = 100$

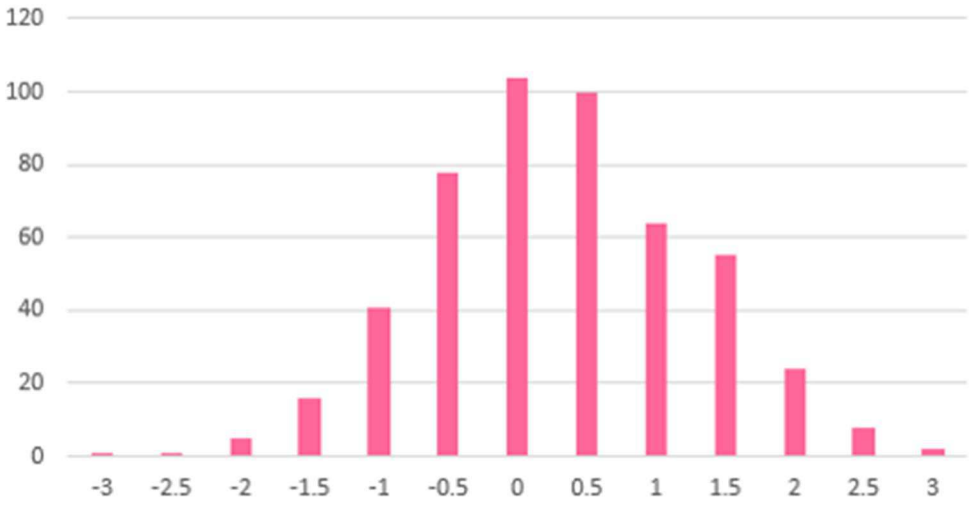

Рисунок 16 – dnorm при $N = 500$

При N = 10 оценка математического ожидания и среднеквадратичного отклонения были равны $\hat{m}$ = 0.11 и $\hat{\sigma}$ = 0.644. При N = 100: $\hat{m}$ = 0.15 и $\hat{\sigma}$ = 0.996. При N = 500: $\hat{m}$ = 0.05 и $\hat{\sigma}$ = 1.010.

В MatLab нормальный закон распределения был реализован при помощи встроенной функции randn и значениях σ = 1, m = 0. Пример использования этой функции представлено на рисунках 17, 18 и 19.

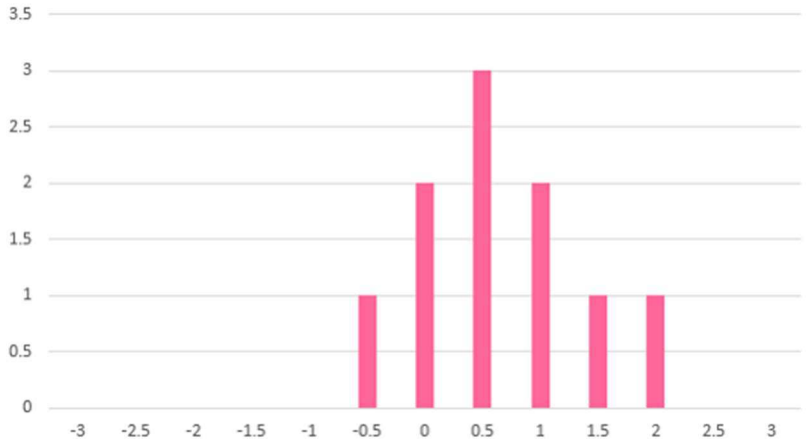

Рисунок 17 – randn при N = 10

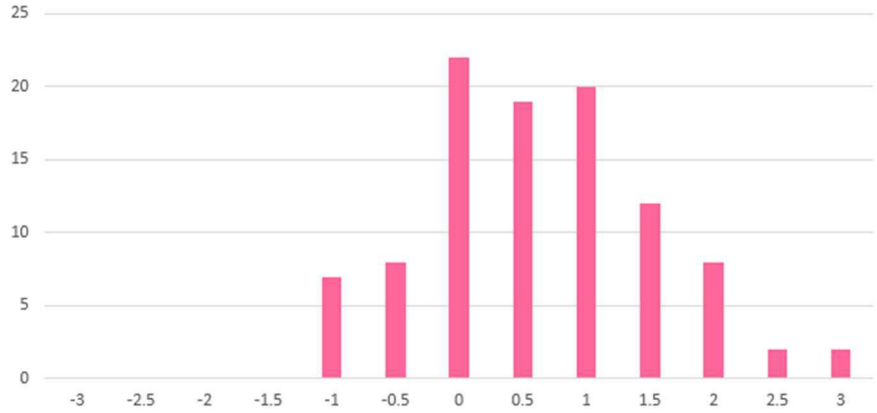

Рисунок 18 – randn при N = 100

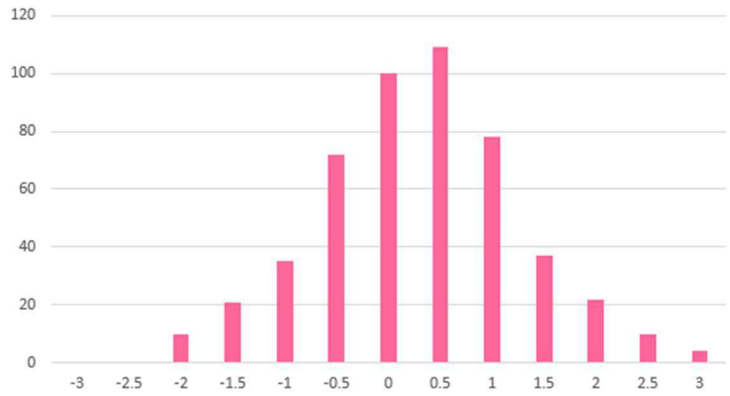

Рисунок 19 – randn при $N = 500$

При $N = 10$ оценка математического ожидания и среднеквадратичного отклонения были равны $\hat{m} = 0.44$ и $\hat{\sigma} = 0.683$. При $N = 100$: $\hat{m} = 0.37$ и $\hat{\sigma} = 0.976$. При $N = 500$: $\hat{m} = 0.04$ и $\hat{\sigma} = 0.987$.

В программном модуле было осуществлено варьирование различных входных параметров.

Ниже, в таблице 1, приведены результаты варьирования значений для нормального закона распределения.

Таблица 1 – Варьирование параметров распределения нормального закона

S	e	k	σ	$\hat{m}$	$\hat{\sigma}$
10	0,005	15	1	0,434	1,584
100	0,005	15	1	0,116	0,989
1000	0,005	15	1	0,015	1,002
1000	0,005	15	1	0,011	0,993
500	0,1	15	1	0,041	0,725
500	0,01	15	1	0,003	0,964
500	0,001	15	1	0,019	1,012
500	0,005	5	1	0,14	1,063
500	0,005	15	1	0,01	0,985
500	0,005	30	1	0,03	0,972
100	0,005	15	1	0,09	0,973
100	0,005	15	3	0,21	2,771
100	0,005	15	10	0,09	8,568

Для наглядности изменения плотности распределения можно взглянуть на гистограмму при варьировании количества подынтервалов и квадратичного отклонения на рисунках 20-25. Черная линия на графике означает заданное значение σ, а красная – ее оценку.

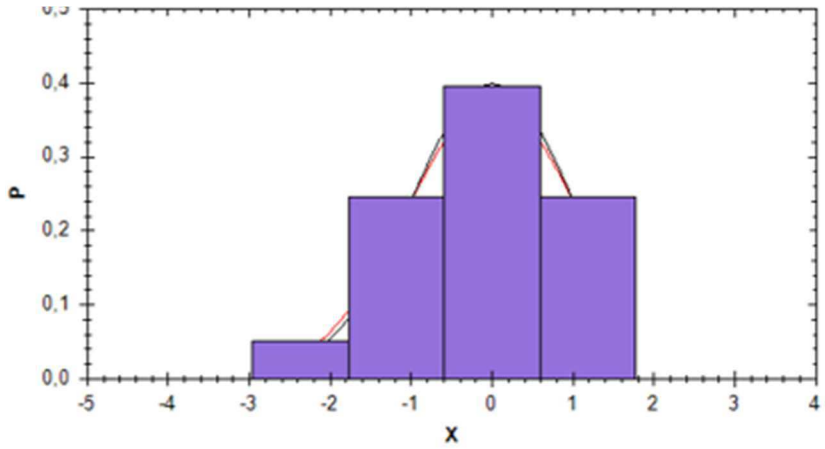

Рисунок 20 – Нормальное распределение при $k = 5$

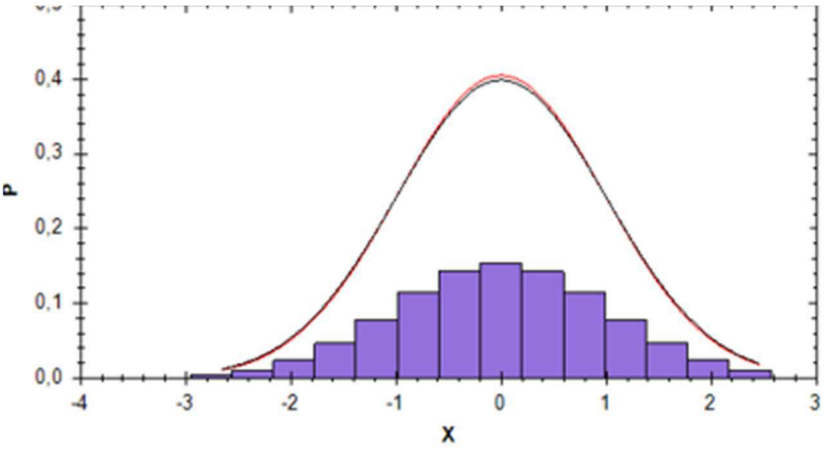

Рисунок 21 – Нормальное распределение при $k = 15$

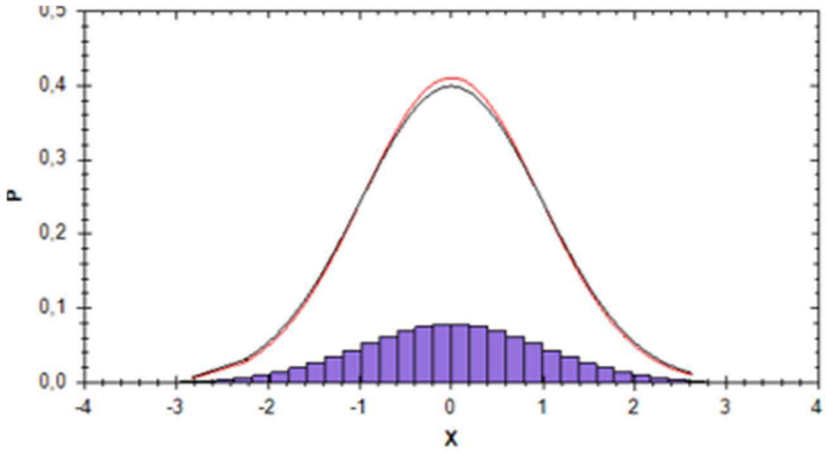

Рисунок 22 – Нормальное распределение при $k = 30$

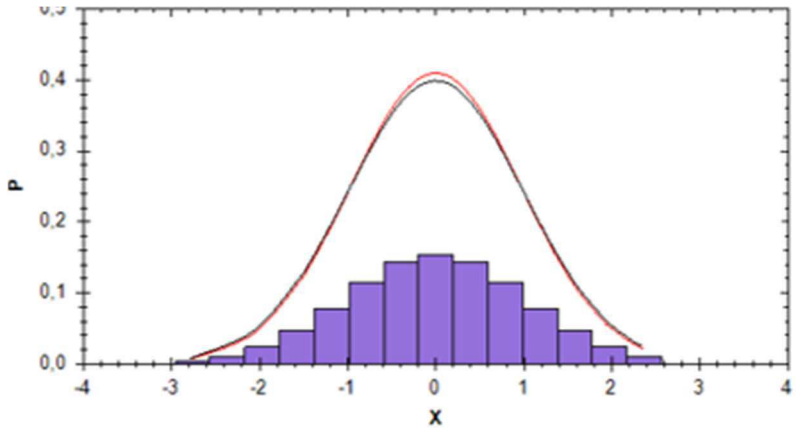

Рисунок 23 – Нормальное распределение при $\sigma = 1$

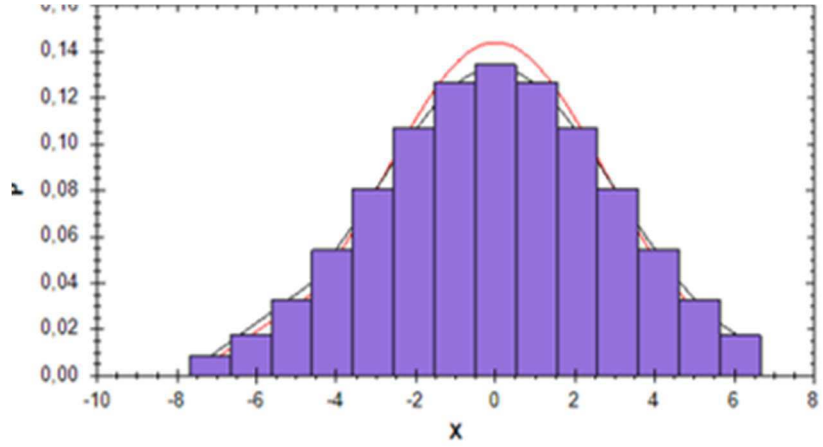

Рисунок 24 – Нормальное распределение при $\sigma = 3$

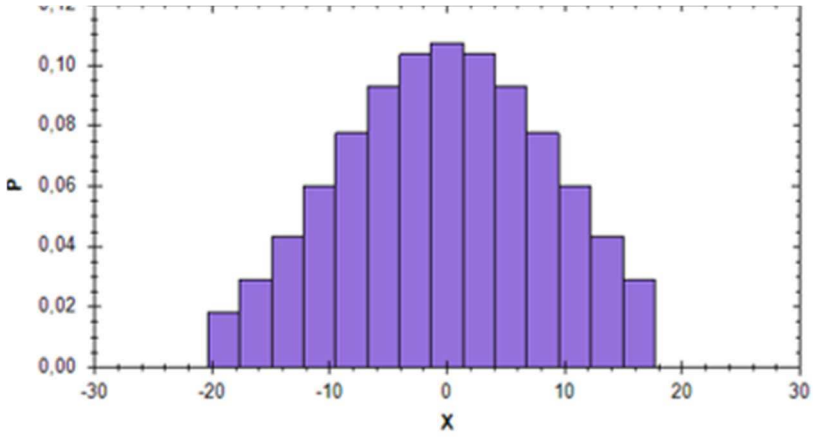

Рисунок 25 – Нормальное распределение при $\sigma = 10$

Результаты варьирования параметров закона распределения Лапласа в виде таблицы и гистограммы при варьировании k и σ приведены ниже, на таблице 2 и рисунках 26-31. Черная линия на графике означает заданное значение λ, а красная – ее оценку.

Таблица 2 – Варьирование значений для закона распределения Лапласа.

S	e	k	λ	$\hat{m}$	$\hat{\lambda}$
10	0,005	15	3	0,04	3,728
100	0,005	15	3	0,003	3,217
1000	0,005	15	3	0,002	2,945
10000	0,005	15	3	0,001	2,969
500	0,1	15	3	0,014	4,091
500	0,01	15	3	0,007	2,673
500	0,001	15	3	0,002	2,996
500	0,005	5	3	0,054	2,237
500	0,005	15	3	0,004	2,898
500	0,005	30	3	0,002	3,525
100	0,005	15	1	0,008	1,148
100	0,005	15	3	0,007	3,149
100	0,005	15	10	0,002	10,506

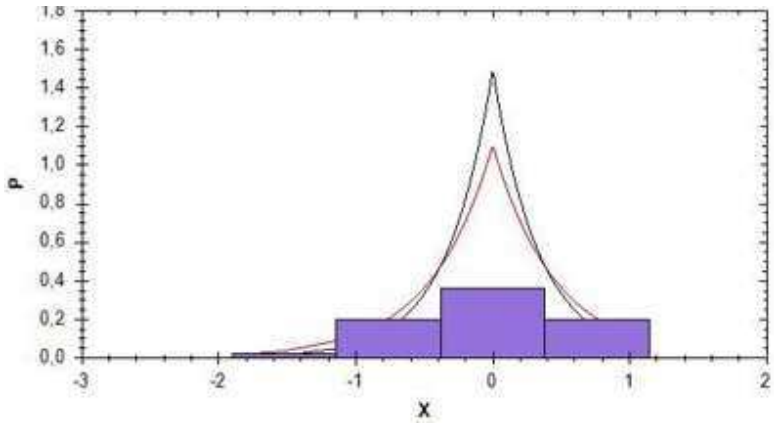

Рисунок 26 – Распределение Лапласа при $k = 5$

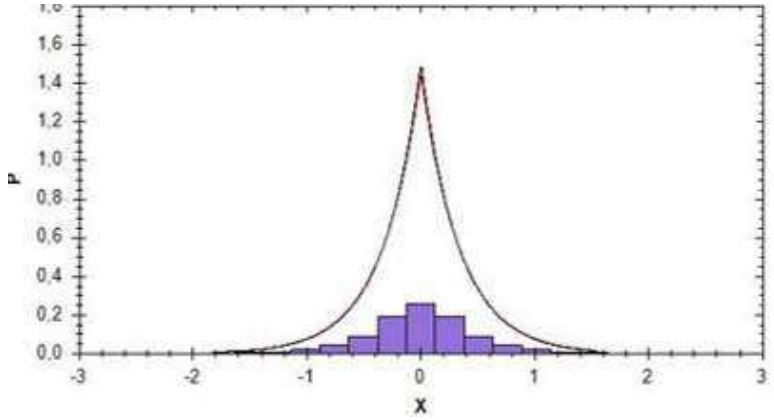

Рисунок 27 – Распределение Лапласа при $k = 15$

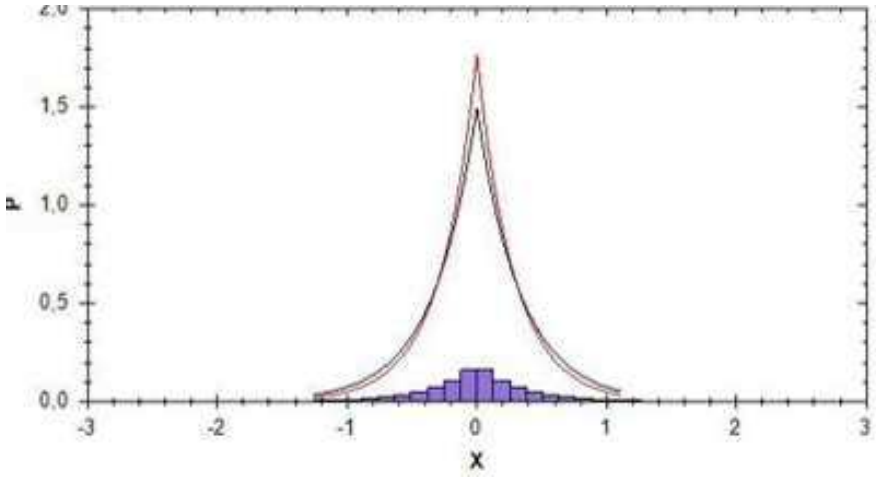

Рисунок 28 – Распределение Лапласа при $k = 30$

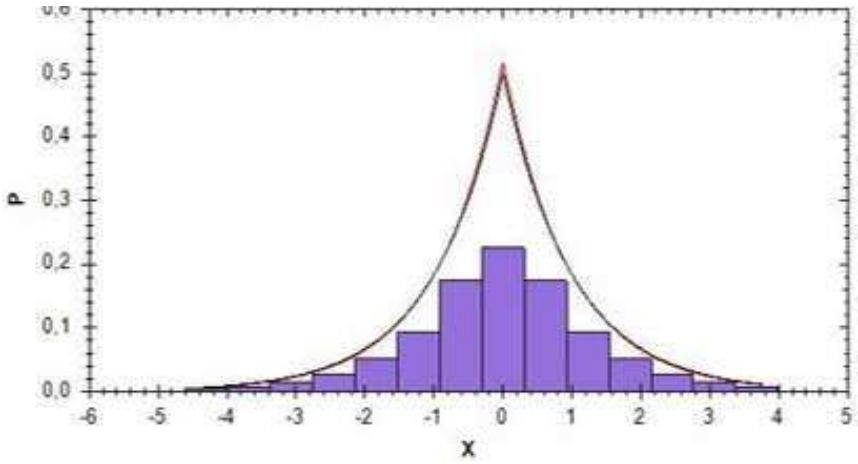

Рисунок 29 – Распределение Лапласа при $\lambda = 1$

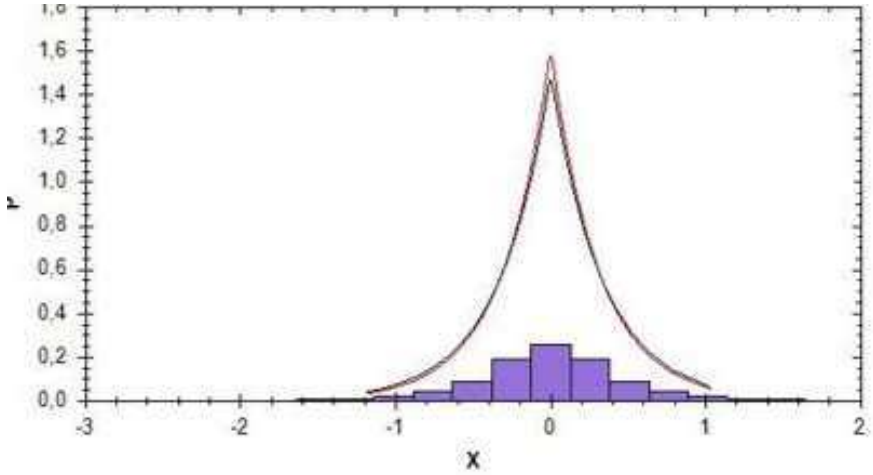

Рисунок 30 – Распределение Лапласа при $\lambda = 3$

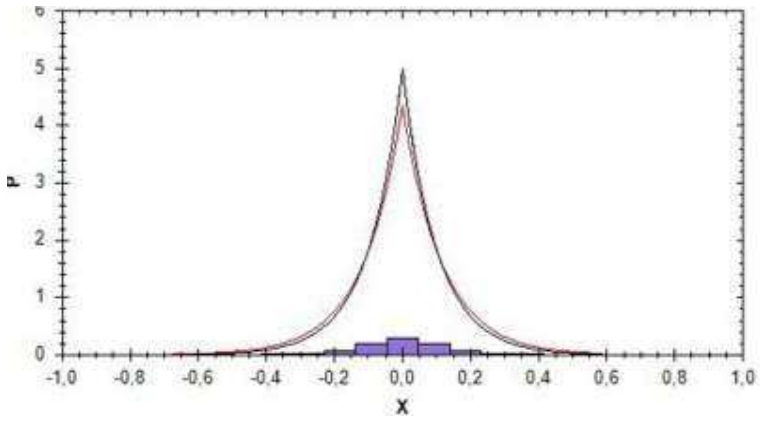

Рисунок 31 – Распределение Лапласа при $\lambda = 10$

Ниже приведены результаты варьирования для закона распределения Парето в виде таблицы и гистограммы при варьировании k и α в таблице 3 и на

рисунках 32-36. Черная линия на графике означает заданное значение α, а красная – ее оценку.

Таблица 3 – Варьирование значений для закона распределения Парето

S	e	k	α	C_0	$\widehat{m}$	$\widehat{\alpha}$	Мин. x
10	0,005	15	2	1	2,501	1,666	1,226
100	0,005	15	2	1	2,04	1,96	1,035
1000	0,005	15	2	1	2,009	1,99	1,005
10000	0,005	15	2	1	2,018	1,981	1,001
500	0,1	15	2	1	1,494	3,022	1,003
500	0,01	15	2	1	1,827	2,208	1,002
500	0,001	15	2	1	1,979	2,021	1,021
1000	0,001	5	2	1	1,579	2,725	1,029
1000	0,001	15	2	1	1,758	2,318	1,009
1000	0,001	30	2	1	2,029	1,972	1,001
500	0,001	15	2	2	3,513	2,321	2,002
500	0,001	15	4	1	1,453	3,50	1,002
500	0,001	15	6	1	1,199	6,026	1,001

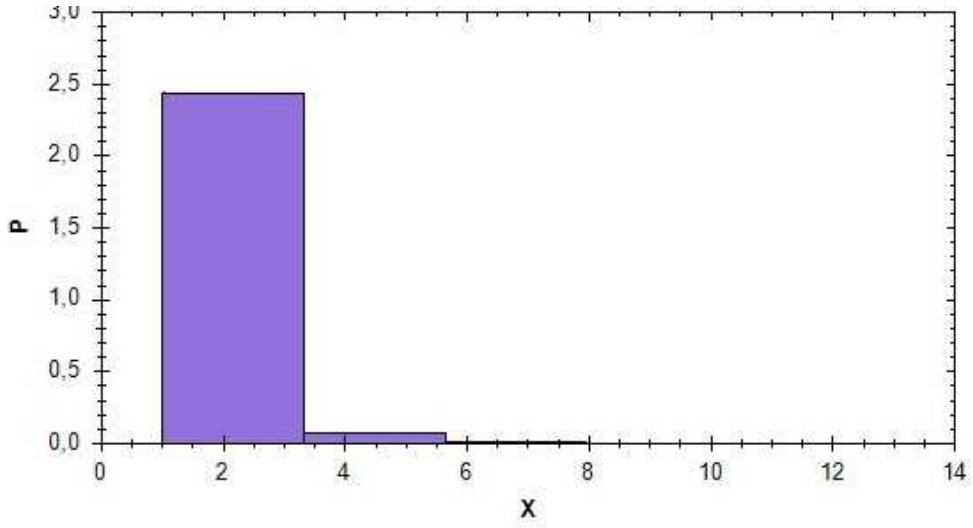

Рисунок 32 – Распределение Парето при $k = 5$

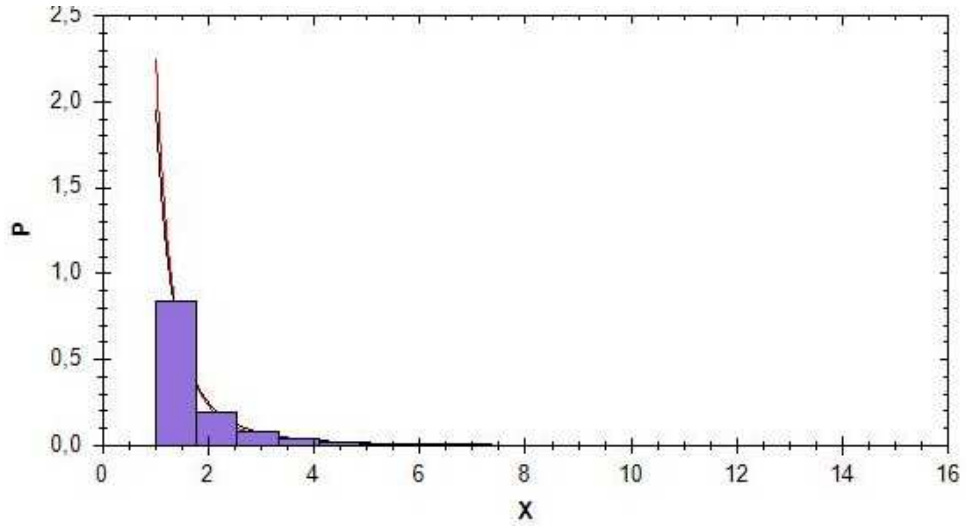

Рисунок 33 – Распределение Парето при $k = 15$

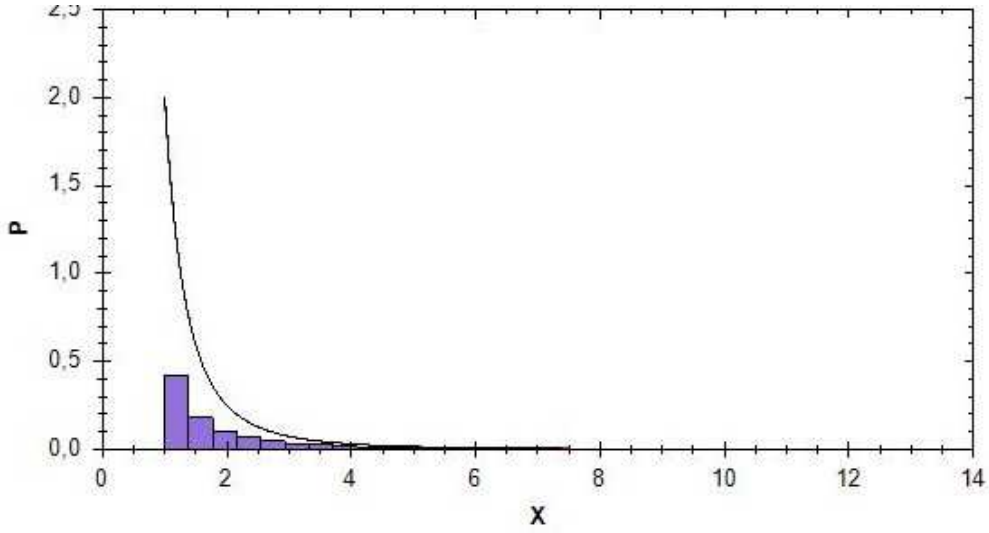

Рисунок 34 – Распределение Парето при $k = 30$

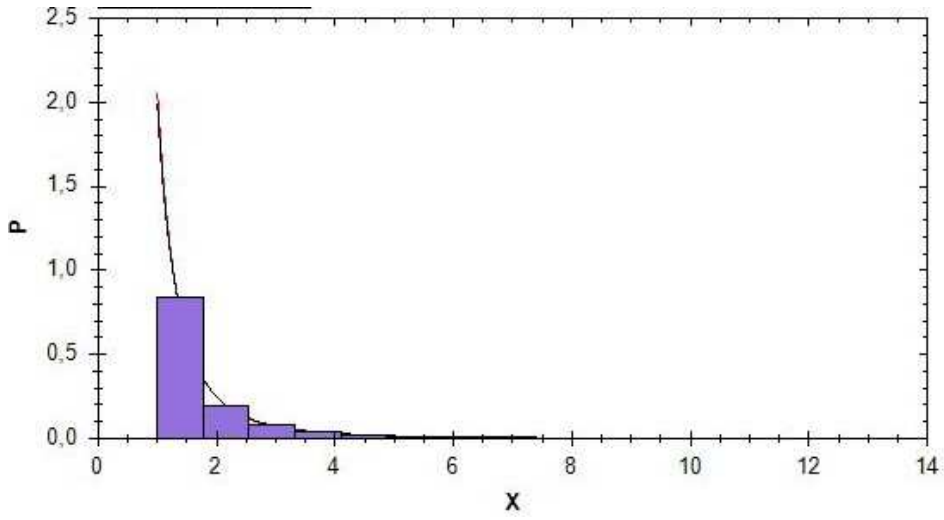

Рисунок 34 – Распределение Парето при $\alpha = 2$

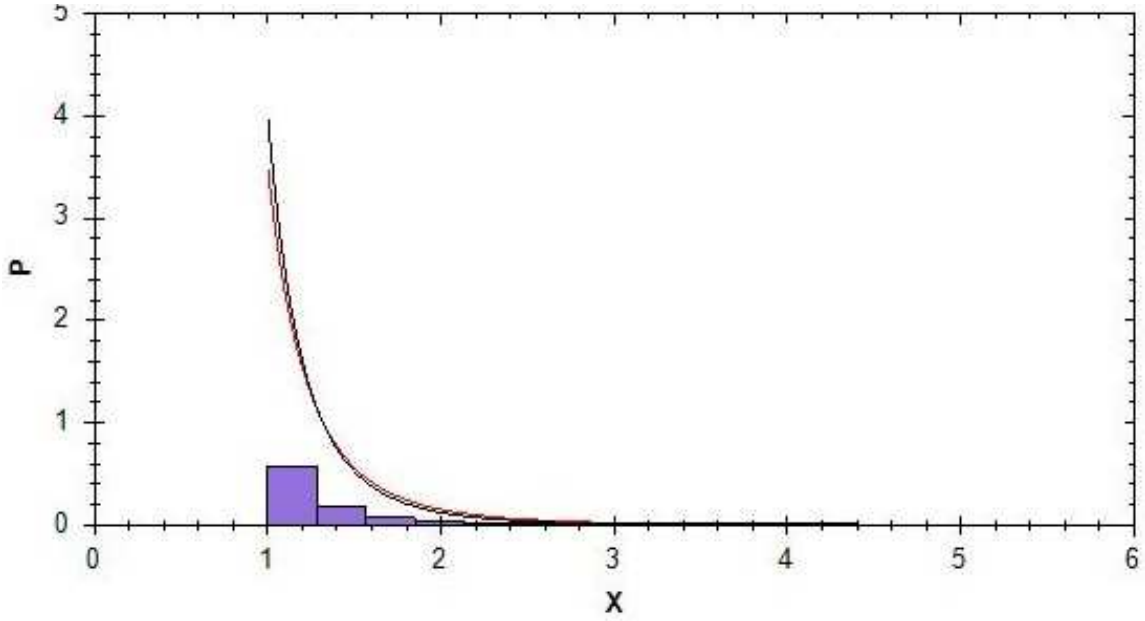

Рисунок 35 – Распределение Парето при α = 4

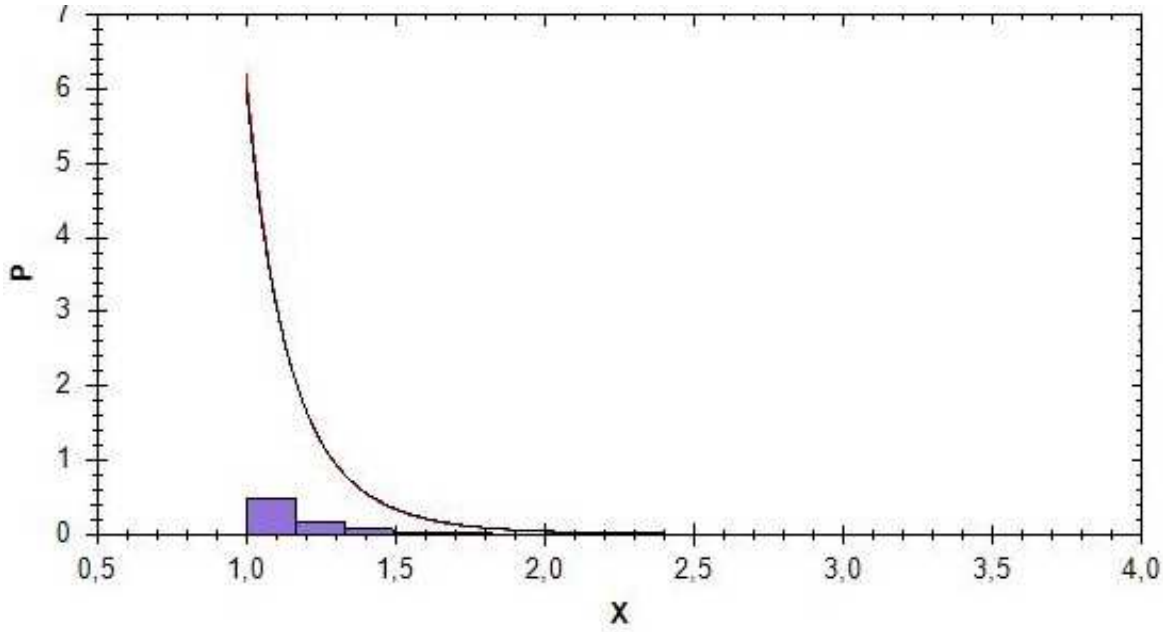

Рисунок 36 – Распределение Парето при α = 6

Ниже приведены результаты варьирования для экспоненциального закона распределения в виде таблицы и гистограммы при варьировании *k* и λ, в таблице 4 и на рисунках 36-42. Черная линия на графике означает заданное значение λ, а красная – ее оценку.

Таблица 4 – Варьирование значений для закона экспоненциального распределения

S	e	k	λ	$\widehat{m}$	$\hat{\lambda}$
10	0,005	15	1	0,791	1,149
100	0,005	15	1	0,869	1,145
1000	0,005	15	1	0,873	1,137
10000	0,005	15	1	1,053	0,939
500	0,1	15	1	0,687	1,452
500	0,01	15	1	0,797	1,25
500	0,001	15	1	0,837	1,191
500	0,005	5	1	0,837	1,032
500	0,005	15	1	0,885	1,126
500	0,005	30	1	0,996	0,999
1000	0,005	15	1	0,868	1,150
1000	0,005	15	3	0,336	2,976
1000	0,005	15	5	0,158	6,331

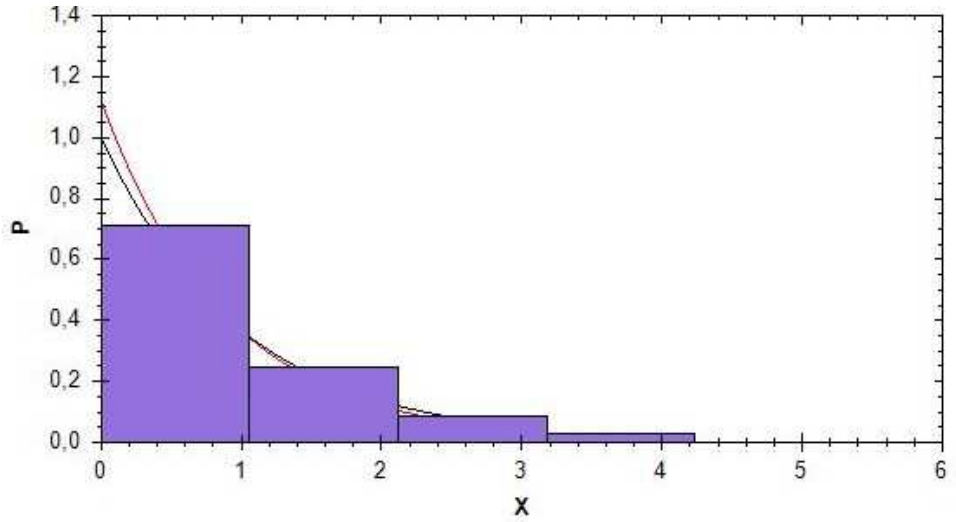

Рисунок 37 – Экспоненциальное распределение при $k = 5$

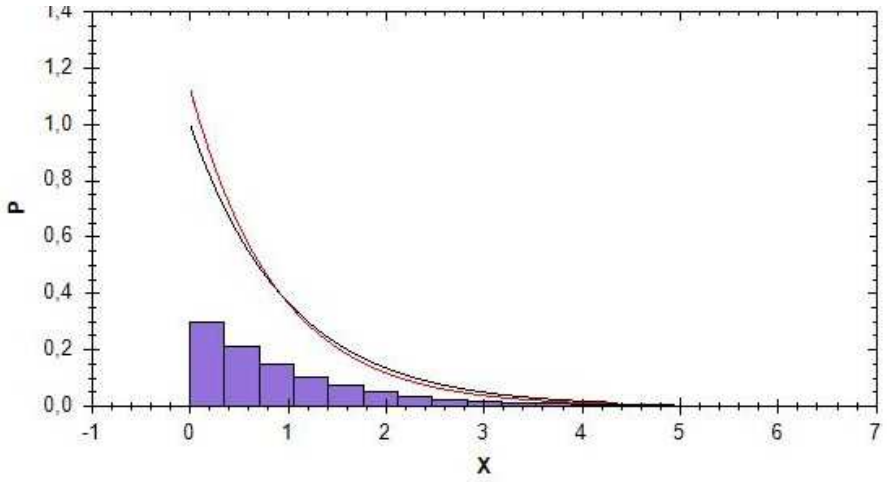

Рисунок 38 – Экспоненциальное распределение при $k = 15$

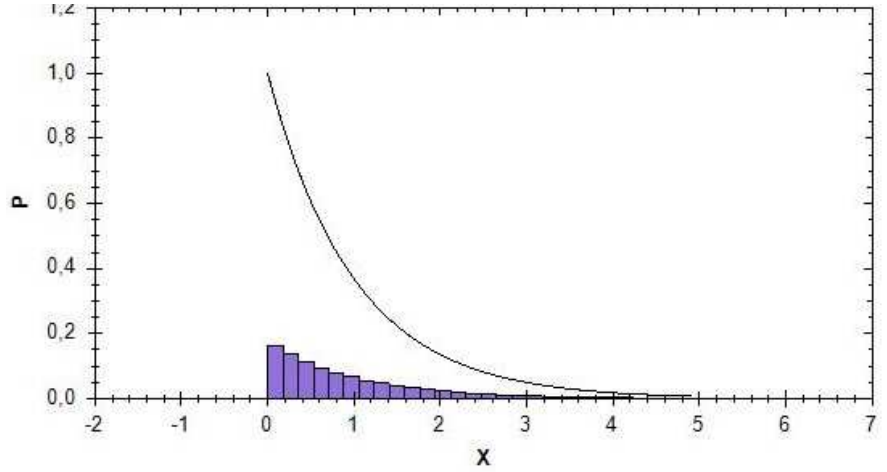

Рисунок 39 – Экспоненциальное распределение при $k = 30$

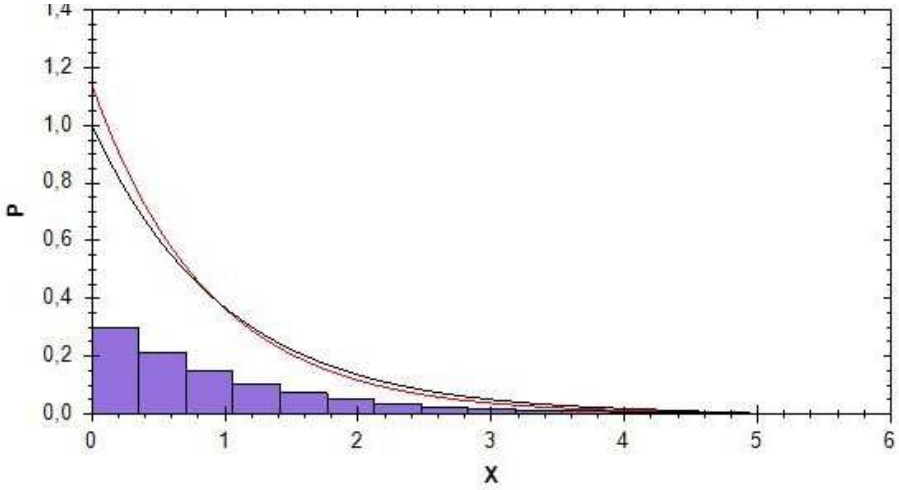

Рисунок 40 – Экспоненциальное распределение при $\lambda = 1$

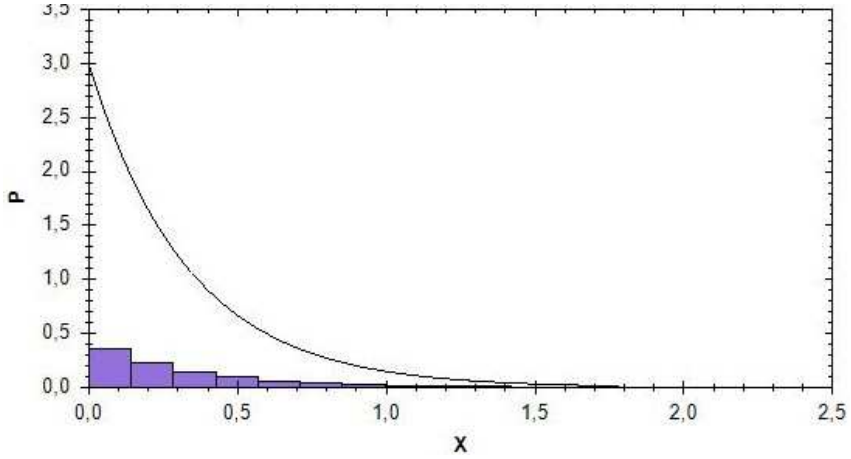

Рисунок 41 – Экспоненциальное распределение при $\lambda = 3$

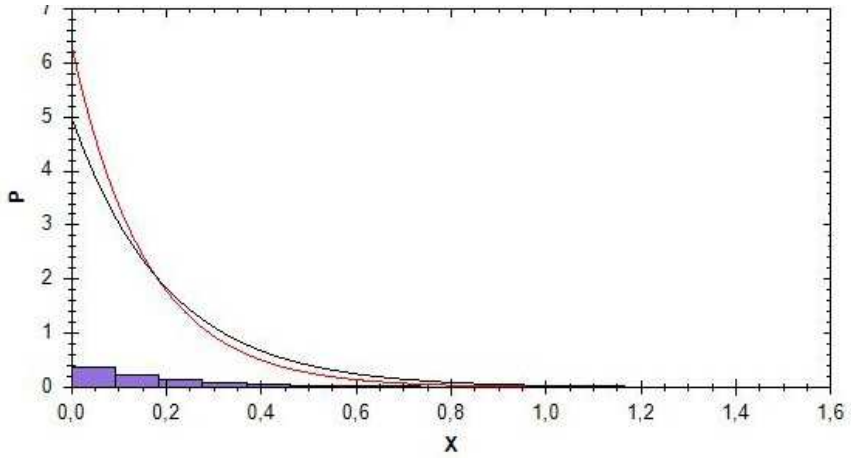

Рисунок 42 – Экспоненциальное распределение при $\lambda = 5$.

3.2 Программный модуль. Руководство пользователя

Разработанный алгоритм генерации псевдослучайных чисел был реализован в виде программного продукта. Главное окно имеет вид, представленный на рисунке 43.

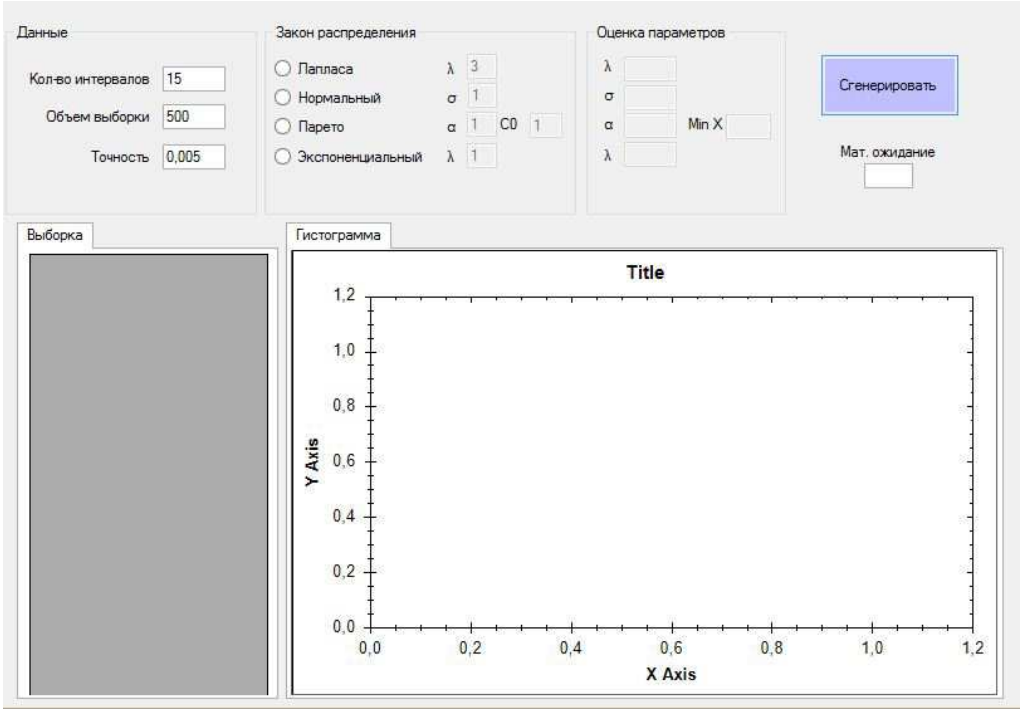

Рисунок 43 – Интерфейс программного модуля

В области «Данные» пользователь задает количество интервалов, объем выборки и точность вычисления. Далее пользователю предлагается выбрать конкретный закон в поле «Закон распределения» и задать параметры распределения напротив выбранного, после чего нажать на кнопку «Сгенерировать».

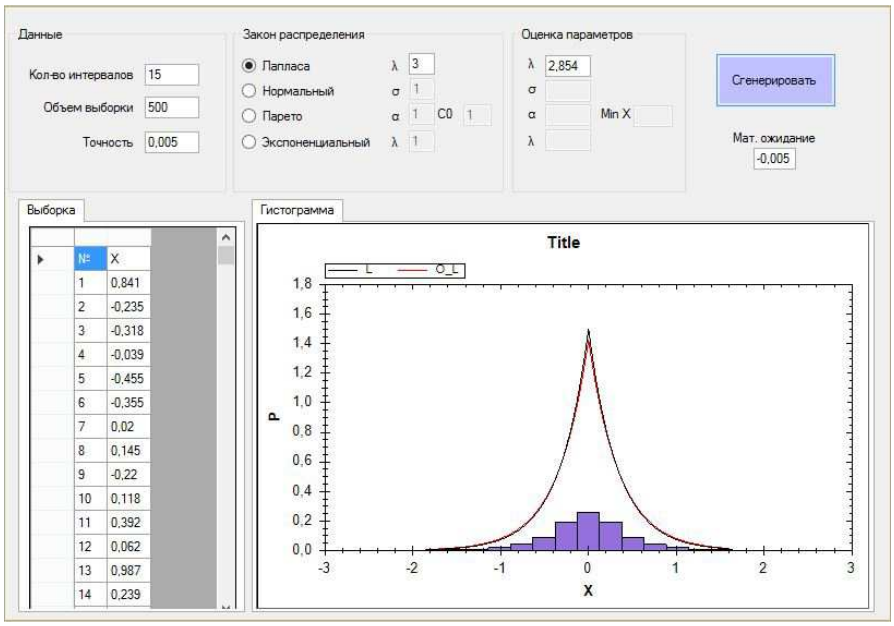

Рисунок 44 – Результат генерации

Сгенерированная выборка представлена в таблице на рисунке 44 (вкладка «Выборка»). График и гистограмма выводятся во вкладку «Гистограмма». В области «Оценка параметров» напротив выбранного закона распределения выводится оценка параметра выбранного распределения. Под кнопкой «Сгенерировать» выводится математическое ожидание закона распределения.

Выборка, сгенерированная программным модулем, записывается в текстовый файл, с названием соответствующего распределения, которым она была сгенерирована, и сохраняется в папку проекта …\bin\Debug. Результат представлен на рисунке 45.

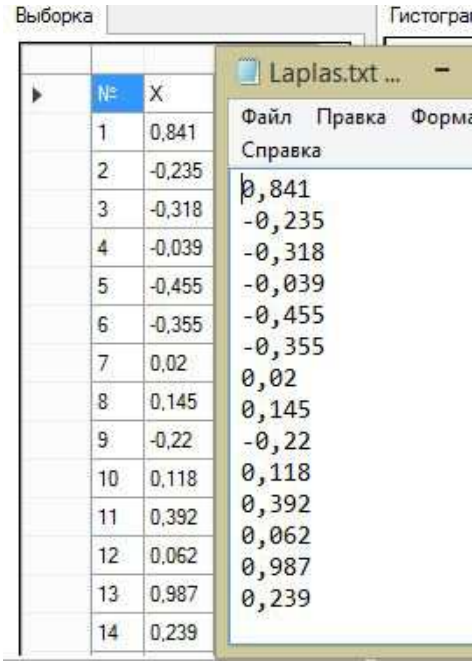

Рисунок 45 – Запись данных в текстовый файл

Выводы по третей главе

Были проведены численные исследования с варьированием значений выборки, количества интервалов, точности и параметров распределения.

По полученным расчётам можно сделать несколько выводов:

– с возрастанием объема выборки расхождение оценки параметра распределения с заданным параметром становится меньше;

– уменьшение точности вычислений уменьшает и разницу между заданным параметром распределения и полученным. Оценка стремится к истинному значению;

– количество интервалов разбиения не оказывает сильного влияния на значения полученных параметров. Это свидетельствует о том, что распределение становится более равномерным.

ЗАКЛЮЧЕНИЕ

Существующие алгоритмы генерации случайных чисел разнообразны. Каждый со своим подходом и применяется для определенных целей. В ходе анализа были выявлены существенные недостатки имеющихся алгоритмов. А также не удалось найти несколько важных и часто используемых на практике законов распределения, реализованных в виде программного модуля.

Был разработан и реализован в виде программного модуля алгоритма на языке C#. Программный модуль работает с четырьмя законами распределения.

По результатам вычислительного эксперимента можно заметить, что разработанный алгоритм имеет ряд преимущества перед существующими. Это видно исходя из более равномерного вида гистограммы и значений оценок параметров распределений, которые стремятся к исходным.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Третьяков, Н. П. Имитационное моделирование методом Монте-Карло и развитие методологии прогнозных оценок макроэкономических показателей / Н. П. Третьяков, Е. О. Щербакова // Технологии техносферной безопасности. – 2009. – № 6.

2. Семенов, А. Д. Идентификация объектов управления: учебное пособие. / А. Д. Семенов, Д. В. Артамонов, А. В. Брюхачев. – Пенза : Изд-во Пенз. гос. ун-та, 2003. - 211 с.

3. Медведев, А. В. Теория непараметрических систем. Управление – I / А. В. Медведев // Вестник Сибирского государственного университета имени академика М. Ф. Решетнева. – 2013. – № 2(48).

4. Спиди, К. Теория управления (идентификация и оптимальное управление). / К. Спиди, Р. Браун, Дж. Гудвин. – Москва : Мир, 1973. – 248 с.

5. Прангишвили, И. В. Идентификация систем и задачи управления: на пути к современным системным методологиям. / И. В. Прангишвили, В. А. Лотоцкий, К. С. Гинсберг // Проблемы управления. – 2004. – № 4.

6. Эйкхофф, П. Основы идентификации систем управления / П. Эйкхофф. – Москва : Мир, 1975.

7. Родионов, И. Б. Теория систем и системный анализ. [Электронный ресурс] : Лекция 9: Классификация видов моделирования систем. – Режим доступа: http://victor-safronov.ru/systems-analysis/lectures/rodionov/08.html.

8. Перегудов, Ф. И. Введение в системный анализ. / Ф. И. Перегудов, Ф. П. Тарасенко. – Москва : Высш. шк., 1989.

9. Рубан, А. И. Методы анализа данных. Учебное пособие / А. И. Рубан // уч. пособие. 2-е изд., исправл. и доп. Красноярск: ИПЦ КГТУ, 2004.

10. Райбман, Н. С. Что такое идентификация? / Н. С. Райбман. – Москва : Наука, 1970.

11. Дилигенская, А. Н. Идентификация объектов управления : учебное пособие. / А. Н. Дилигенская. – Самара : Самар. гос. техн. ун-т., 2009. – 9-21 с.

12. Сергеева, Н. А. П-генератор случайных чисел по закону Лапласа. / Н. А. Сергеева, М. В. Цепкова, Е. А. Чжан // Решетневкие чтения. – 2011. – №15.

13. Цыпкин, Я. З. Информационная теория идентификации / Я. З. Цыпкин. – Москва : Наука, 1995.

14. Медведев, А. В. Теория непараметрических систем. Моделирование / А. В. Медведев // Вестник Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. – 2010. – № 4.

15. Медведев, А. В. Теория непараметрических систем. Процессы / А. В. Медведев // Вестник Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. – 2010. – №3.

16. Медведев, А. В. Теория непараметрических систем. Общий подход / А. В. Медведев // Вестник Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. – 2008. – №3.

17. Тарасенко, Ф. П. Моделирование и феномен человека. Ч. 1 / Ф. П. Тарасенко. – Москва : Науч. технологии, 2012.

18. Цыпкин, Я. З. Основы теории атематических систем. / Я. З. Цыпкин. – Москва : Наука, 1977.

19. Антонов, А. В. Системный анализ. Учеб, для вузов / А. В. Антонов. – Москва : Высш. шк., 2004.

20. Соболь, И. М. Метод Монте-Карло. / И. М. Соболь. – Москва : Наука, 1968.

21. Первушин, В. Ф. Прецизионный генератор случайных чисел. / В. Ф. Первушин, Н. А. Сергеева, А. В. Стрельников // Вестник Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. – 2010. – № 5.

22. Соболь, И. М. Численные методы Монте-Карло. / И. М. Соболь. – Москва : Наука, 1973.

23. Гроп, Д. Методы идентификации систем. / Д. Гроп. – Москва : Мир, 1979. – 302 с.

24. Бендат, Дж. Измерение и анализ случайных процессов. / Дж. Бендат, А. Пирсол. – Москва : Мир, 1974. – 463 с.

25. Морданев, Е. Е. Понятие о статистическом моделировании систем массового обслуживания / Е. Е. Морданев // Новая наука: проблемы и перспективы. – 2017. – № 2.

26. Рубан, А. И. Теория вероятностей и математическая статистика. Учебно-методическое пособие / А. И. Рубан. – Красноярск: Сиб. федер. ун-т, 2012.

27. Прохоров, Ю. В. Вероятность и математическая статистика: энциклопедия / Ю. В. Прохоров. – Москва : Большая Рос. энцикл., 2003.

28. Ермаков, С. М. Курс статистического моделирования. / С. М. Ермаков, Г. А. Михайлов. – Москва: Наука, 1976.

29. Гмурман, В. Е. Руководство к решению задач по теории вероятностей и математической статистике : учебное пособие / В. Е. Гмурман. – Москва : Высш. школа, 2002.

Федеральное государственное автономное образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Институт космических и информационных технологий
Базовая кафедра интеллектуальных систем управления

УТВЕРЖДАЮ
Заведующий кафедрой
_____ Ю. Ю. Якунин
«11» июня 2018 г.

БАКАЛАВРСКАЯ РАБОТА

27.03.03 Системный анализ и управление

Разработка алгоритма генерации случайных чисел

Руководитель 11.06.18 ст. преподаватель Е. А. Чжан
подпись, дата должность, ученая степень инициалы, фамилия

Выпускник _____ А. С. Тихонов
подпись, дата инициалы, фамилия

Красноярск 2018